Д. М. Сегал

ОСНОВЫ *фонологической* СТАТИСТИКИ

Издательство «Наука»

АКАДЕМИЯ НАУК СССР
ИНСТИТУТ СЛАВЯНОВЕДЕНИЯ И БАЛКАНИСТИКИ

Д. М. Сегал

ОСНОВЫ ФОНОЛОГИЧЕСКОЙ СТАТИСТИКИ

(на материале польского языка)

Издательство «Наука»
МОСКВА 1972

В монографии сформулирована основная проблема фонологической статистики — проблема однородности текста относительно частот фонологических элементов. В работе содержится критико-исторический обзор теоретических положений фонологической статистики за последние 50 лет, дается очерк фонологии польского языка, на материале которого решалась проблема.

Ответственный редактор

доктор филологических наук
И. И. РЕВЗИН

7-1-1
376—72

ПРЕДИСЛОВИЕ

Настоящая работа выполнялась в течение нескольких лет в секторе структурной типологии славянских языков Института славяноведения АН СССР. Первоначально задуманная как статистическое описание польского языка на парадигматическом и синтагматическом уровнях, эта книга претерпела ряд изменений в процессе работы. Прежде всего, часть, посвященная синтагматической статистике, оказалась гораздо менее полной, чем раздел о статистике в парадигматике. Произошло это главным образом в результате того, что автор счел необходимым основное внимание уделить вопросу, который представляется ему кардинальным для лингвистической статистики вообще — проверке однородности текстов относительно частот лингвистических объектов.

Таким образом, работа приняла характер обсуждения одной проблемы. Представляется, впрочем, что именно подробное и углубленное изучение одной проблемы — как раз то, чего зачастую недостает лингвостатистике, где более распространены обширные рассуждения по целому кругу проблем одновременно.

Автор считал также необходимым включить в работу подробное обсуждение лингвостатистической литературы, особенно трудов основоположника лингвистической статистики Дж. К. Ципфа и одного из крупнейших современных специалистов в этой области, недавно скончавшегося английского ученого Г. Хердана.

Это, по нашему мнению, должно продемонстрировать важное значение проблемы однородности в фонологической статистике. Разумеется, в одной работе нельзя было затронуть все вопросы, связанные с применением статистических методов в фонологии, равно как и соответствующую литературу; однако автор решил сознательно ограничиться узким кругом проблем, что и обусловило выбор обсуждаемой литературы. Мы надеемся, что поставленные вопросы станут предметом научного обсуждения.

Автор искренне благодарит канд. физ.-мат. наук Ю. И. Левина, чьи критические замечания во многом помогли формированию концепции книги, и Е. М. Сегал, которая оказала неоценимую помощь в подготовке рукописи к печати и без чьей моральной поддержки эта работа не была бы закончена.

ВВЕДЕНИЕ

Цель настоящей работы — подвергнуть критическому анализу основные положения, на которых базируется лингвистическая статистика фонологического уровня, а также дать соответствующее статистическое описание фонологии польского языка.

К настоящему времени число работ, в которых содержится статистика фонологических элементов, настолько велико, а попытки теоретического осмысления и экспериментальной проверки основных положений (зачастую выраженных неявно), делающих подобную статистику осмысленной, настолько редки, что представлялось необходимым рассмотреть основы фонологической статистики как лингвистической дисциплины.

Три фактора, делающие возможным статистическое рассмотрение лингвистических данных,— это массовость языковых высказываний, повторяемость языковых объектов в этих высказываниях и случайность выпадения данного элемента.

Указанные три фактора характеризуют любые статистические системы событий, поэтому естественно, что им должны удовлетворять и лингвистические совокупности, рассматриваемые как статистические. Обычно считается, что все три фактора действительно характерны для лингвистических совокупностей, сомнение возникает лишь при рассмотрении лингвистических событий как случайных. Однако это сомнение, как правило, снимается указанием на то, что сознательный выбор, явно имеющий место при употреблении языковых объектов, всегда выступает «в паре» со случайностью, что случайность фигурирует также в тех аспектах (например, стилистических), которые обычно относятся к сфере «выбора» и, наконец, что употребление языкового объекта вызывается столь сложным комплексом причин, что результирующая может быть приравнена к действию случайности.

Таким образом, обеспечивается законность применения статистики на «прагматическом» уровне.

Представляется, однако, что проблема гораздо сложнее. Она состоит не в том, чтобы доказывать, что л ю б а я языковая совокупность является статистической, а в том, чтобы найти в языковой совокупности такой слой, который действительно является статистическим. При этом доказательством «статистичности» будет не априорное утверждение о том, что языковая совокупность отвечает тем или иным требованиям, предъявляемым к статистической совокупности, а реальный анализ наблюденных фактов с точки зрения статистики.

Одним из наиболее трудных вопросов применения статистических методов к лингвистике является вопрос о релевантных единицах, т. е. о соотношении статистической и лингвистической релевантности при выделении единицы подсчета (так называемых учетных единиц). Естественно стремление к тому, чтобы учетные единицы были лингвистически релевантными, однако обычные языковые единицы (фонемы, слова, морфемы) выделяются на основании внестатистических соображений. Задача состоит в том, чтобы определить, являются ли подобные единицы статистическими, и найти действительно статистические единицы, если языковые единицы таковыми не окажутся. Далее предстоит интерпретировать обнаруженные статистические единицы лингвистически.

Статистическими могут быть признаны единицы, образующие статистические совокупности, характеризуемые, во-первых, вышеуказанными тремя факторами. Далее, статистические совокупности должны быть однородными, т. е. входящие в них элементы должны оставаться на всем протяжении совокупности качественно идентичными, причем таким образом, чтобы произведенная случайным образом выборка из любой части совокупности давала совершенно адекватное представление о всей совокупности. Иными словами, основные параметры двух произвольных выборок из одной совокупности должны совпасть.

Как соотносятся статистические понятия совокупности и выборки с понятиями, выработанными для описания языка? Следует указать, что в лингвистической статистике этот вопрос принадлежит к числу наименее разработанных. Применительно к фонологическому уровню обычно можно встретить утверждение, что статистической генеральной совокупностью здесь являются все тексты на данном языке, что подобная совокупность однородна и что, следовательно, любой отдельный текст будет являться случайной выборкой из этой совокупности. Применительно к лексическому уровню такое утверждение также иногда имеет место, однако здесь можно встретить и рассуждение о том, что совокупностью является не «весь язык» как бесконечный текст, а его более узкая сфера (например, функциональный стиль),

достаточно большая группа текстов (произведения художественной литературы определенной эпохи), произведения одного автора и т. д. вплоть до отдельного текста. С другой стороны, не текст, а словарь может пониматься как совокупность.

Не говоря уже о том, что подобные расхождения свидетельствуют о полной теоретической и практической неразработанности проблемы, они указывают на фундаментальную сложность языка как статистического объекта. Эта сложность приводит к тому, что положения, принципиально существенные для статистического осмысления языка (например, о бесконечности генеральной языковой совокупности), заменяются прямо противоположными (текст конечен) с тем, чтобы ближе подойти к лингвистической реальности.

В настоящей работе, в частности, будут рассмотрены вопросы о характере текста как случайной выборки на фонологическом уровне.

Статистические единицы, образующие однородные совокупности, обладают, по определению, некоторой априорной характеристикой — в е р о я т н о с т ь ю. Именно эта характеристика и полагается постоянной для данной совокупности. Иными словами, при статистическом подходе совокупность предполагается однородной относительно вероятности. Здесь мы подходим к основному исходному пункту всей лингвистической статистики. Этим пунктом является постулирование существования априорной теоретической вероятности для данной совокупности. Соответственно предполагается очевидным, что априорная вероятность на уровне наблюдения представлена ч а с т о т о й конкретных языковых объектов (фонем, морфем, слов) и что эта частота статистически адекватно представляет вероятность. Предполагается, что отличия частоты от априорной вероятности, наблюдаемые на опыте, не превышают колебаний, возникающих под действием чистой случайности. Такова общепринятая в лингвистической статистике схема образования наблюденных частот языковых объектов.

На практике однородность языковых совокупностей должна была бы означать, что наблюденные в реальных текстах частоты одинаковых элементов (особенно фонем, так как фонологический уровень наименее зависит от сознательного выбора) достаточно стабильны, чтобы при надлежащей статистической проверке их можно было возвести к общему прототипу.

Однако в действительности оказывается, что подобная картина неверна. Существующие статистические процедуры не дают возможности найти общий прототип для большинства реально наблюденных частот лингвистических объектов, относящихся к самым различным языковым уровням. Соответственно исходный постулат лингвистической статистики о том, что для каждого лингвистического объекта существует некоторая априорная теоретическая вероятность, приходится поставить под сомнение. Дело не

только в том, что существующие статистические процедуры неадекватны (зачастую различие между наблюдаемыми частотами одного и того же элемента настолько велико, что не требует специальных процедур для своего установления), но в том, что реальные наблюденные частоты языковых объектов, по-видимому, являются результатом работы нескольких весьма различных механизмов, а не только одного чисто статистического механизма. Задача настоящего исследования в том, чтобы попытаться частично вскрыть эти механизмы.

При самом первом рассмотрении фактов можно заметить по крайней мере две противоположные тенденции в поведении частот языковых элементов. Эти две тенденции отражают работу двух различных механизмов. С одной стороны, существует интуитивное осознание того, что языковым элементам (или, говоря осторожнее, некоторым из них) присуща постоянная величина встречаемости (частость или редкость), которая как бы абстрагируется, отделяется от конкретных манифестаций и становится постоянной характеристикой элемента. Тот факт, что определенная встречаемость может быть постоянной характеристикой элемента и, следовательно, должна быть учтена в лингвистическом описании, можно проиллюстрировать на частном примере существующих систем в фонологии [1]. Если традиционно отношения в системе фонем понимались как отношения между в общем равноправными элементами, то в упомянутых фонологических работах имеется тенденция учитывать особое положение, занимаемое маргинальными элементами, которые помимо дистрибутивных, или внутренних, характеристик (ср. носовое [ã] в немецком языке в слове Chance при отсутствии дифференциального элемента «носовость» в системе гласных) выделяются и своей крайней редкостью.

Определенные языковые элементы могут быть охарактеризованы как частые (например, гласный [а] во многих языках). При этом частость подобных элементов не может быть существенно изменена даже путем сознательного отбора: частота элемента of в английском языке весьма велика и не зависит от выбора ввиду очень большой роли этого элемента в грамматической структуре английского языка.

---

[1] Упомянем здесь специально полонистическую работу: T. Milewski. Derywacja fonologiczna. «Biuletyn Polskiego Towarzystwa Językoznawczego», t. IX (1949). В общем плане, а также специально на материале чешского языка эту проблему поднял Кучера (H. Kučera. Inquiry into co-existing phonemic systems in Slavic languages. 's-Gravenhage, 1958. Два доклада о сосуществовании в фонологической системе разноплановых элементов были сделаны на V Международном конгрессе фонетических наук в Мюнстере: H. Pilch. Zentrale und periphere Lautsysteme. «Proceedings of the 5th International Congress of Phonetic Sciences». Basel — N. Y., 1965; J. Vachek. On peripheral phonemes. Там же.

Таким образом, для некоторых языковых элементов частость или редкость можно считать такой характеристикой, которая релевантна не только для одного определенного текста, но может быть распространена и шире.

С другой стороны, столь же заметны случаи, когда один и тот же элемент гораздо чаще употребляется в одних текстах, чем в других. Для лексического уровня это самоочевидно, однако и для фонологического уровня можно представить себе ситуацию, в которой определенная фонема будет в некоторых текстах встречаться заведомо чаще, чем в других, например, в русских текстах по философии фонема [ф] будет наверняка встречаться чаще, чем в текстах обиходной речи.

Чтобы быть лингвистически значимым, статистическое описание языка должно отражать работу этих двух противоположных тенденций, а не сводиться к простому перечню частот. Можно предположить, что тенденция к стабильной частоте отражает функционирование чисто статистического «слоя» языковой совокупности, в то время как тенденция к нестабильной частоте отражает действие текстовых, внестатистических механизмов. Тот факт, что обе тенденции проявляются в пределах одного языкового уровня, показывает, что неверно априори считать последовательность элементов данного уровня статистической совокупностью. Надлежит обнаружить чисто статистические элементы среди языковых объектов. В нашем случае эта задача будет решаться для фонологического уровня.

Поскольку обнаружение структуры в плане статистической организации языка соотносится с изучением структуры языка вообще, встает вопрос о месте статистического описания в лингвистике. В лингвистической литературе вопрос о соотношении статистического и структурного описания дебатировался довольно широко, при этом представители структурного направления обычно считали, что статистическое описание не затрагивает язык в соссюровском понимании.

В частности, Н. С. Трубецкой полагал, что данные, полученные на разных текстах, будут настолько сильно отличаться друг от друга, что не имеет смысла говорить о средних значениях встречаемости элемента или его качества (например, длительности гласного). Язык — система отношений и поэтому «...лежит вне меры и числа»[2]. Еще более категоричен в этом смысле Є. К. Шаумян: «Фундаментом структурной лингвистики служит принцип иерархии, а иерархия по своей природе антистатична, так как она является детерминистской. Отсюда все изложенное подтверждает известный тезис о том, что методы структурной

---

[2] Н. С. Трубецкой. Основы фонологии. М., 1960, стр. 15.

лингвистики должны быть не вероятностные, а детерминистские»[3].

В общем, тезис о том, что статистический аспект чужд языку, характерен для течений структурной лингвистики, подчеркивающих чисто реляционную природу языка. При таком понимании вся вероятностная сторона языковых явлений неизбежно считалась принадлежащей речи.

Данная проблема не может быть должным образом поставлена до тех пор, пока не будут выяснены основные, т. е. статистические, вопросы. Сначала следует определить, насколько постоянной является частость или редкость элементов; интерпретация может быть дана лишь после того, как сам факт постоянства будет точно установлен.

Общетеоретические споры, особенно о плане того, лежит или не лежит язык «вне меры и числа», кажутся бесплодными. Возможны по крайней мере два способа их урегулирования (в случае, если гипотеза постоянства окажется справедливой):

— признать, что факт постоянства встречаемости относится к сфере языка. Подобное решение будет уместным, если в язык будет включен и инвентарь языковых элементов, как это имеет место в дескриптивной лингвистике, а также в теории моделей языка;

— признать, что факт постоянства встречаемости не относится к сфере языка. Подобное решение отвечает реляционной концепции языка. В этом случае, однако, придется как-то решать вопрос о соотношении средней встречаемости, характерной для элемента вообще, с его конкретной встречаемостью в данном тексте. Что в таком случае следует считать речью — конкретные манифестации (дающие конкретную величину встречаемости) или некоторую «речь в целом», для которой характерна средняя (статистически постоянная) встречаемость? В результате образуется не дихотомия «язык — речь», а трихотомия «язык — Речь — речь».

В настоящей работе предлагается подход к данной проблеме, отвлекающийся от онтологии языка, которая требует, чтобы мы определили отношение статистического описания к природе языка («является ли язык строго детерминистской системой или допускает и вероятностный подход», «совместимы ли статистические методы с реляционной природой языка»). В основе этого подхода лежит отношение к лингвистическому описанию как к моделированию. Сущность лингвистического моделирования состоит в том, что основной акцент переносится на выяснение не того, чем язык является, а того, как он работает. Аналогию такому подходу можно найти в разнообразных приложениях естественных наук, когда природа многих систем остается неясной, при том, что известно, как они функционируют.

---

[3] С. К. Шаумян. Структурная лингвистика. М., 1965, стр. 335.

За последнее время появилось много работ, посвященных рассмотрению лингвистического моделирования [4]. В своей последней книге «Метод моделирования и типология славянских языков» И. И. Ревзин так определяет метод моделирования: «Метод моделирования ...состоит в построении модели, т. е. некоторой системы знаков (логическое моделирование) или же некоторой системы физических объектов (физическое моделирование), обладающих следующими свойствами:

1) исходные данные модели (в кибернетике их называют «входом») соответствуют некоторой существенной части совокупности исходных объектов;

2) модель действует таким образом, что результат ее действий (в кибернетике его называют «выходом») соответствует некоторой существенной части совокупности заключительных объектов.

Описанное здесь понимание моделирования принято, например, в работах по машинному переводу и вполне согласуется с тем, как предлагает применять идеи моделирования к языку И. А. Мельчук» [5]. Вот что пишет по данному вопросу И. А. Мельчук:

«Не останавливаясь специально на сложном теоретическом вопросе о сущности научного описания, мы можем считать, что весьма эффективным способом создания и проверки описаний каких-либо систем является построение действующих моделей этих систем. Поясним, что имеется в виду.

Предположим, что мы рассматриваем совокупность каких-либо объектов, порождаемых скрытым от нас механизмом. Нас интересует как раз этот механизм, но он недоступен непосредственному наблюдению и о нем можно судить только по результатам его деятельности, т. е. по свойствам совокупности объектов, порождаемой этим механизмом. При этом мы интересуемся данным механизмом в строго определенном отношении: нам важно знать лишь о тех сторонах его функционирования, которые обусловливают порождение им рассматриваемой совокупности. Никакие конкретные особенности механизма и его функционирования для нас не существенны.

---

[4] Из советской литературы см., в частности: А. А. Зиновьев, И. И. Ревзин. Логическая модель как средство научного исследования. «Вопросы философии», 1960, № 1; И. И. Ревзин. Модели языка. М., 1962; Он же. Некоторые вопросы теории моделей языка. «Научно-техническая информация», 1964, № 8; В. Вс. Иванов. О применимости фонологических моделей. «Труды ИТМ и ВТ АН СССР», вып. 2. М., 1961; С. К. Шаумян. Проблемы теоретической фонологии. М., 1962; Он же. Структурная лингвистика; О. С. Ахманова, И. А. Мельчук, Е. В. Падучева, Р. М. Фрумкина. О точных методах исследования языка. М., 1961.

[5] И. И. Ревзин. Метод моделирования и типология славянских языков. М., 1967, стр. 25.

Анализируя данную нам совокупность объектов, порожденных механизмом, мы создаем гипотетическое описание этого механизма. Чтобы проверить наше описание, можно построить на его основе модель механизма, так как очень много конкретных свойств механизма не будет учтено, и в некоторых отношениях модель вовсе не будет похожа на самый механизм. Но если эта модель, функционируя, будет порождать в точности те же самые объекты, что и исследуемый механизм, то можно считать, что в интересующем нас отношении наша модель адекватна и что, следовательно, наше описание верно» [6].

Таким образом, в порождающих моделях выходом должен явиться тем или иным способом упорядоченный корпус лингвистических объектов.

И. И. Ревзин выделяет два типа порождающих моделей: «Модель первого типа (например, синтез) имеет на входе описание определенного факта внешнего мира, т. е. совокупность десигнатов и общий смысл высказывания, а на выходе — определенную фразу, т. е. последовательность слов. Если взять n таких моделей, то мы получим на выходе конечное число фраз, но мы не можем, по-видимому, получить бесконечное множество фраз, ибо для этого надо было бы уметь зафиксировать бесконечное число фактов внешнего мира.

Модель второго типа (например, порождающая грамматика) и отличается тем, что она при конечном входе порождает бесконечное множество фраз. Это достигается тем, что эта модель абстрагирована от реального содержания высказываний и учитывает лишь некоторые общие смысловые категории» [7].

Как в модели первого, так и в модели второго типа порождаются реальные фразы, поэтому в них должен быть предусмотрен словарь морфем (следовательно, и фонем), с помощью которого производится идентификация определенных цепочек отношений со «знаконосителями» (sign vehicles) [8]. Словарь этот может мыслиться и вне модели, но тогда в модели должно быть устройство, с помощью которого производится выбор из словаря.

Очевидно, что для целей синтеза модель должна быть снабжена вероятностными характеристиками лингвистических элементов на разных уровнях. Введение таких характеристик является одним из способов выбора между несколькими вариантами грамматических или лексических морфов в словаре. Более того, представляется, что синтезированные тексты будут «естественными» только в том случае, если в них будет соблюдено то же соотношение

---

[6] О. С. Ахманова, И. А. Мельчук, Е. В. Падучева, Р. М. Фрумкина. Указ. соч., стр. 40—41.

[7] И. И. Ревзин. Метод моделирования и типология славянских языков, стр. 27.

[8] См., в частности: M. Halle. The Sound Pattern of Russian. 's-Gravenhage, 1959.

редких и частых элементов, что и в моделируемом объекте. Соответственно модель синтеза должна строиться на основе анализа и имитирования реальных языковых текстов. Статистические показатели в таких моделях выступают в роли операторов, осуществляющих перевод модели из статистического в динамическое состояние. Важную роль играют статистические показатели в организации словарей элементов — более часто встречающиеся элементы должны, естественно, быть более доступными. В отношении к моделям порождения вопрос обстоит сложнее. В теории порождающих грамматик существует направление [9], полагающее, что смысл работы порождающей грамматики состоит в «рекурсивном перечислении правильно построенных предложений», т. е. цель грамматики — установление правильности, а реальные операции выбора морфем лежат за ее пределами. «Порождающая грамматика сама по себе не синтезирует и не анализирует предложения, она нейтральна по отношению к говорящему и слушающему» [10].

Соответственно все вопросы, связанные со списком элементов и т. п., считаются не относящимися к модели. Нам кажется, что не существует резкого противопоставления между перечислением и «производством». Часто оказывается, что тип допустимой трансформации, правильность конструкции могут зависеть от типа лексической морфемы, поэтому мы полностью соглашаемся со следующим положением И. И. Ревзина: «Хомский подчеркнул, что построенные таким образом порождающие модели не совпадают с моделью говорящего и, соответственно, обратные им модели не совпадают с моделью слушающего, и это верно в том смысле, что логически возможна ситуация, когда перечислить все фразы языка удобнее одним способом, в то время как каждая фраза в действительном акте общения будет производиться иначе. Однако такая ситуация кажется маловероятной. Скорее всего в речевой деятельности человека одни и те же механизмы ответственны и за перечисление всех фраз, и за производство каждой конкретной фразы» [11].

Более того, нам представляется, что существующие порождающие грамматики включают статистические характеристики (правда, в неявном виде). Выделение особого набора ядерных конструкций, путем трансформации которых порождаются остальные предложения, может опираться, помимо прочих критериев (обычно нигде не постулируемых), и на большую их частость.

---

[9] См., например: M. Chomsky. On the notion «rule of grammar». «Proceedings of symposia in applied mathematics», v. XI. Structure of language and its mathematical aspects. N. Y., 1961.

[10] Р. Б. Лиз. О возможностях проверки лингвистических положений.— ВЯ, 1962, № 4; С. К. Шаумян. Структурная лингвистика.

[11] И. И. Ревзин. Метод моделирования и типология славянских языков, стр. 27.

В существующих моделях порождения вероятностные критерии могут быть не только применимы к словарям морфем, но, в конечном итоге, они могут быть использованы для разделения некоторых типов грамматической синонимии. Речь идет о таких известных случаях синтаксических синонимов, как *агрессивная политика — политика агрессии*, *шел лесом* — и *шел через лес*, *говорить страстно* и *говорить со страстью*, а также о морфологических вариантах типа *цеха — цехи* или англ. two mink — two minks.

С. К. Шаумян полагает эти варианты равноправными интерпретациями одной фразы [12]. На уровне аппликативной модели эти парные варианты можно считать полностью равноправными, так как они сопоставляются с одинаковыми цепочками отношений. Однако, если считать, что конечная цель модели порождения — значимый выбор элементов и порождение реального текста, то на каком-то уровне требуется этот выбор осуществить, и здесь решающими могут оказаться статистические критерии.

Таким образом, в порождающих моделях необходимы статистические параметры, которые бы были применимы для организации словаря модели, а также задавали бы порядок работы правил.

Эти правила могут иметь простой вид: «Выбрать наиболее частый элемент», или же они могут быть даны в виде различных импликаций типа «Если А, то выбрать самый частый элемент», или «Если В, то выбрать самый редкий элемент» и т. д.

Все, что было сказано о необходимости включения статистических характеристик в нестатистические модели порождения и синтеза, объясняет необходимость практической работы по нахождению таких характеристик. Однако собственно статистические исследования языка имеют и свою теоретическую ценность. В статистических исследованиях язык рассматривается как динамический процесс, имеющий свои закономерности. Эти закономерности, с одной стороны, интересны в плане изучения структуры языковых высказываний, бо́льших, чем предложение (направление, именно теперь выдвигающееся на первый план в лингвистических исследованиях),— как раз в таких высказываниях структурность может быть обнаружена в первую очередь в результате статистических наблюдений (ср. работы Г. А. Лесскиса). С другой стороны, статистические характеристики в высшей степени необходимы при построении модели языковой коммуникации, при рассмотрении языка в аспекте кодирования и проч. В этом плане языковая деятельность как организованный процесс несомненно имеет четкие статистические закономерности, которые могут оказаться чрезвычайно важными при изучении связей между различными уровнями языка (например, проблема длины слова, с одной

---

[12] С. К. Шаумян. Структурная лингвистика, стр. 334.

стороны, в связи с морфологической структурой языка, а с другой стороны, в связи с распределением частот для слов). Соответственно статистическая сторона языкового процесса также может и должна быть предметом моделирования.

Все сказанное о методе моделирования как о способе понять функционирование объекта, который невозможно описать непосредственно, относится и к представлению языка как статистической совокупности. Выше мы упоминали о том, что картина реальных частот языковых объектов, получаемая при анализе текстов, сложна, противоречива и не описывается простой схемой, согласно которой эта частота репрезентирует априорную вероятность. Наша задача — создать такую модель, которая бы объясняла возникновение этих реальных частот.

Соответственно, можно считать, что каждый текст характеризуется определенной картиной соотношения частот языковых объектов (распределением частот), которая синтезируется в речевом процессе аналогично синтезу самих объектов. Наша модель должна (хотя бы приблизительно) объяснить, как синтезируется распределение частот на данном уровне. При этом наше объяснение должно быть релевантным не только для данного текста и данного распределения, но и для других текстов и распределений.

В статистике существует целый ряд априорных теоретических распределений (нормальное распределение, распределение Пуассона), позволяющих предсказывать вид фактических частот, если наблюдаемое распределение подходит под указанное теоретическое. Отсюда интерес, который лингвистическая статистика проявляет к нахождению теоретического распределения языковых объектов, которое бы адекватно подходило к экспериментальным данным.

В настоящей работе будут затронуты вопросы применимости теоретических распределений статистики к языковым объектам, здесь мы укажем, что, поскольку априорные распределения заранее предполагают статистическую однородность совокупности, построение модели образования реальных частот языковых объектов должно опираться прежде всего на экспериментальную проверку предположения о статистической однородности языковых совокупностей.

Поскольку постулат об однородности текстов относительно частот лингвистических единиц является центральным постулатом всей лингвистической статистики, и при этом до сих пор он принимался без экспериментальной проверки, автор счел необходимым выполнить настоящую работу прежде всего как проверку положения об однородности на фонологическом уровне. Описание статистических характеристик польской фонологии будет производиться под углом зрения стабильности или нестабильности частот отдельных элементов.

Таким образом, задача настоящей работы — проверка некоторых основных постулатов фонологической статистики и описание в этом плане статистических характеристик польской фонологии — будет выполняться в виде построения модели, объясняющей синтез реальных частот. Модель эта должна будет включать как часть, объясняющую наблюдаемую стабильность частот, так и часть, ответственную за их нестабильность. Построение обеих частей модели будет производиться в процессе проверки однородности текстов по существующим статистическим критериям.

Глава первая

ЛИНГВИСТИЧЕСКАЯ СТАТИСТИКА И ФОНОЛОГИЯ

Практические потребности обращения с языком привели к тому, что количественный аспект лингвистических явлений стал изучаться сравнительно давно. Пожалуй, одной из древнейших областей прикладного языкознания является криптография — наука о тайнописи, шифрах и способах их дешифровки. Именно в пределах криптографии впервые было обращено внимание на определенную стабильность, постоянство встречаемости букв (а, следовательно, и звуков) в данном языке. Практически знания подобного рода используются уже много столетий, а теоретически они получили обоснование в XIX в. [1] Количество букв сравнительно невелико, их комбинации перечислимы, поэтому легко было заметить количественные закономерности в поведении букв в текстах.

С развитием техники письма и распространением средств массовой коммуникации количественная сторона речевой деятельности становится достоянием не только криптографов, но и всех, кто занимается созданием множительных устройств и систем письма. Проблема выявления наиболее частых букв и их сочетаний с целью экономного и рационального устройства клавиатур пишущих и типографских машин, а также для наиболее оптимальной организации систем стенографии уже давно занимает специалистов в соответствующих областях. Для этого еще в XIX в. предпринимались подсчеты частоты букв и их сочетаний [2]. Следует

[1] Ср., например, руководство: A. Kerckhoffs. La cryptographie militaire. Paris, 1883, где приводится частотность букв и их сочетаний во французском языке в сравнении с другими языками.

[2] См. для французского языка: F. Dujardin. Journal des conaissances usuelles. Paris, 1834, а также: J. J. Thierry-Meg. Structure phonologique du Française. Phonographie à peinte unique. Nouveau système d'écriture abrégée. Paris, 1813. Для немецкого языка — знаменитый словарь: F. K ä-

отметить, что приведенные в сноске работы Дюжардена и Тьери-Мег (не говоря уже о словаре Кединга), а также стенографические подсчеты звуков и букв, выполненные для других языков, долго сохраняли свою значимость в качестве источников фактического материала о частотности языковых элементов. В частности, Дж. К. Ципф черпал некоторые из своих данных именно из работы Дюжардена — спустя почти сто лет после ее опубликования.

Работы подобного рода, т. е. подсчеты частотности звуков (фонем), выполненные для практических целей, и составляют один тип исследований по лингвистической статистике применительно к плану выражения. Таких работ вышло очень много и не всегда их можно учесть, так как зачастую они появляются в крайне специализированных и недоступных изданиях [3]. Не все они могут быть использованы даже в качестве источников материалов, так как иногда авторы не различают в них букв и звуков (не упоминая о различении звуков и фонем). Однако за последние годы появилось несколько работ по составлению систем алфавита и стенографии для языков развивающихся стран. Некоторые из этих описаний выполнены на блестящем лингвистическом и техническом уровне — например, работы по статистическому описанию новоиндийских языков. Такова книга д-ра Ш. В. Бхагват [4], который обследовал 100 000 слов (553 860 фонем) языка маратхи и привел статистику слов, слогов, морфем и фонем наряду со статистикой алфавитных элементов. По охвату материала этот подсчет превосходит многое из того, что известно для других языков. Во всяком случае, русский язык еще ждет такого детального обследования. С точки зрения лингвистической и теоретико-статистической работа д-ра Бхагват абсолютно непредубежденна и ясна. Как нам стало известно, книга д-ра Бхагват представляет собой часть обширного проекта по статистическому описанию языков Индии, предпринятого Декканским колледжем университета в Пуне, университетами штатов Майсор, Гуджарат и др. Помимо маратхи такие описания уже выполнены для хинди, гуджарати и каннада, ведется работа по другим языкам.

---

d i n g. Häufigkeitswörterbuch der deutschen Sprache. Steiglitz b. Berlin, 1898.

[3] Ср., например, работу: J. S e d l á č e k. Základní studii k českému těsnopisu. S. Štanoveni poměrů frekvenčních, iteračních a kombinačních v jazyce českem. Как указано в библиографии P. G u i r a u d. Bibliographie critique de la statistique linguistique. Utrecht. Anvers, 1954, рукопись этой работы находится в Государственном институте стенографии в Праге, а ее изложение помещено в специальном стенографическом журнале «Těsnopisné Rozhledy» (Praha, 1924).

[4] Shriram Vasudeo B h a g w a t. Phonemic Frequencies in Marathi and Their Relation to Devising a Speed-Script. Poona, 1961. Статистическое описание языка гуджарати также вышло из печати. См.: Pandit Prabodh B e c h a r d a s. Phonemic and Morphemic Frequencies of the Gujarati Language. Poona, 1965.

Трудно переоценить значение подобных описаний. Они предоставляют незаменимый материал не только для практических целей, но и для целей лингвистической типологии. К сожалению, подсчеты такого рода пока исчисляются единицами. Их основное достоинство — огромный охват материала — как раз то, что отличает эти работы от многочисленных работ по подсчету частот фонем, выполненных за последние 40 лет и не преследующих прикладных целей. Эти подсчеты фонем выполняются в чисто описательных целях и иногда по нескольку раз для одного языка.

Для примера можно привести данные из статьи Ванг и Крофорд [5], где сравниваются результаты подсчетов частоты английских согласных фонем, опубликованные начиная с 1874 по 1959 г. Таких подсчетов, по данным Ванг и Крофорд, было десять (следует добавить, что после выхода в свет обзора Ванг и Крофорд появилось одиннадцатое исследование статистики английских фонем [6], из которого мы и будем заимствовать наши данные в дальнейшем). Часть из них была выполнена в чисто описательных целях, другие в прикладных. Авторы обзора приходят к следующим выводам:

1. Существует значимая разница между результатами подсчетов частот фонем по словарю и по тексту. Этот вывод вполне естествен и послужил основой для фонологической статистики Пражского лингвистического кружка.

2. Различия в выборе инвентаря для представления фонологического уровня могут привести к тому, что результаты подсчетов частот окажутся несравнимыми и релевантными лишь для данной системы фонологической транскрипции.

Этот вывод имеет большое теоретическое значение, так как вскрывает первичность фонологического описания по отношению к статистике — объекты подсчета должны быть фонологически значимыми, и их надо выбирать с учетом возможных сопоставлений.

3. Относительная частота согласных существенно не зависит от стиля или содержания текста, а также от его размера: наблюдается хорошая корреляция между результатами, полученными еще в 1874 г. У. Д. Уитни [7] на тексте объемом в 10 000 фонологических элементов, и результатами К. Вёлькера [8] — общий объем текста 409 506 элементов.

Этот вывод является практически очень существенным — появляется возможность использования малых выборок для статисти-

---

[5] W. S.-Y. Wang, J. Crawford. Frequency studies of English Consonants. «Language and Speech», 1960, v. 3, p. 3, стр. 131—139.

[6] P. B. Denes. On the statistics of spoken English. «Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung». Berlin, 1964, Bd 17, H. 1, стр. 51—72.

[7] W. D. Whitney. The Proportional elements of English utterance. «Proocedings of the American Philological Association», 1874, v. 14.

[8] C. H. Voelker. A comparative study of investigations of phonetic dispersion in connected American speech. «Archives néerlandaises de phonetique expérimentale», 1937, v. 13.

ческих фонологических подсчетов. Однако, как нам представляется, его нельзя трактовать так, что любой текст даст величины частотности, совпадающие с результатами, полученными на крупной выборке. Результаты на малых выборках могут совпасть с результатами на больших выборках, однако вероятность такого совпадения будет увеличиваться по мере увеличения выборки. Поэтому то, что для английского языка проведено так много фонемных статистик, помогает оценить результаты каждого конкретного подсчета — каждый эксперимент можно рассматривать как контрольный для другого.

В тех же случаях, когда статистика основывается на единичной (притом малой) выборке, ее с трудом (если вообще) можно признать значимой [9].

К сожалению, в отличие от английского, для подавляющего большинства языков, подвергнутых статистической обработке, подсчеты выполнены на малых единичных выборках, и их нельзя считать достоверными. Поэтому возможности типологии, основанной на фонологической статистике, пока ограничены небольшим количеством достоверных источников.

* * *

В принципе все работы по применению статистических методов к фонологическому уровню языка можно разделить на три группы по возможности в них, соответственно, фонологических и статистических аспектов.

Первая (наиболее многочисленная) группа — это уже упоминавшиеся прикладные и чисто описательные подсчеты. Вторая группа — это работы, использующие статистику для выяснения чисто фонологических вопросов, и третья, включающая в себя меньше всего работ, посвящена решению чисто статистических проблем.

Первые работы, посвященные сравнению различных языков с точки зрения частоты встречаемости фонологических элементов, появились в середине XIX в. [10], а первым полным статистическим описанием фонологической системы языка явились работы замечательного американского лингвиста У. Д. Уитни. Выше уже говорилось о том, что его статистическое описание английского языка выполнено так, что сохраняет свою ценность и значимость

---

[9] Ср., например: Séan de B ú r c a. Irish phoneme frequencies, «Orbis», t. IX 1960, № 2.

[10] E. F ö r s t e m a n n. Numerische Lautverhältnisse im Griechischen, Lateinischen und Deutschen. «Zeitschrift für vergleichende Sprachforschung begr. von A. Kuhn», Bd I. Göttingen, 1852, стр. 163—173; О н ж е. Numerisch Lautbeziehungen des Griechischen, Lateinischen und Deutschen zum Sanskrit. Там же, Bd 2, 1853, стр. 35—44; О н ж е. Numerische Lautverhältnisse in Griechischen Dialekten. Там же, стр. 401 и сл.

по сей день. То же можно сказать и о его статистическом описании системы фонем санскрита [11]. Уитни раскрыл неодинаковый удельный вес различных элементов в системе и выразил это различие в терминах встречаемости в тексте. Его подсчеты как по английскому языку, так и по санскриту основываются на выборках по 10 000 элементов из связного текста. Это — первая попытка дать описание отношений в парадигматике, основываясь на синтагматических данных.

Вслед за появлением первых подсчетов частоты встречаемости число работ по количественному анализу стало увеличиваться, и в конце XIX в. вышел в свет уже упоминавшийся первый частотный словарь, составленный Ф. Кедингом для немецкого языка. Совершенно естественным явилось стремление дать количественным фактам какую-то интерпретацию. Поскольку тогдашнее общее языкознание интересовалось исключительно проблемами генетического родства языков, а проблемы описания языка как системы выдвинулись на первый план позднее, интерпретация пришла со стороны внелингвистических естественных наук, и прежде всего позитивной психологии. Психолого-биологическая трактовка явлений встречаемости в языке утвердилась на несколько десятков лет. Из работ такого рода следует упомянуть книгу Б. Бурдона [12]. Влияние этой книги, с одной стороны, на квантитативное языкознание и, с другой стороны, на психологию языка было достаточно значительным. Бурдон приводит таблицы относительной частоты звуковых элементов во французском, немецком, итальянском, испанском, русском, английском, венгерском языках. Эти данные впоследствии неоднократно использовались в лингвистической статистике (Ципф, Хердан). К сожалению, как это сознавал и Ципф, пользовавшийся таблицами Бурдона, его данные фактически недостоверны. Правда, для целей Ципфа этим можно было пренебречь (см. об этом ниже), однако некоторые лингвисты, прибегавшие к работе Бурдона, этого не знали, так как черпали материал уже из вторых рук. Дело в том, что Бурдон основывал свои подсчеты на слишком малых выборках (не больше 3 000 элементов), что, как мы покажем далее, не обеспечивает даже однократной встречаемости всех элементов.

Кроме того, он брал тексты либо в алфавитной записи, либо в записи, крайне произвольно отражающей фонологическую систему языка. Это в сочетании с недостаточным объемом материала привело автора к весьма произвольным выводам. Например,

---

[11] W. D. Whitney. On the comparative frequency of occurence of the alphabetic elements in Sanskrit. «Oriental and Linguistic Studies», 2nd series. N. Y., 1874; Он же. Sanskrit Grammar. Boston, 1896.

[12] B. Bourdon. L'éxpression des émotions et des tendences dans le langage. Paris, 1892; M. Weiss. Über die relative Häufigkeit der Phoneme des Schwedischen. «Statistical Methods in Linguistics», 1961, № 1, стр. 45 и сл. (в этой статье используются результаты Бурдона).

среди девяти самых частых согласных английского языка отсутствует [ð], хотя по данным всех позднейших фонетических подсчетов для английского языка этот согласный занимает примерно пятое — шестое место среди самых частых согласных. Эти и другие ошибки подрывают доверие к данным Бурдона. Однако книга, вышедшая почти восемьдесят лет назад, нуждается в снисхождении. Его, однако, не заслуживают те лингвисты, которые апеллируют к так называемому закону Бурдона о безусловном преобладании во всех языках дентальных звуков (ср. статью М. Вайсс о шведской фоностатистике)[13]. Конечно, данные Бурдона опосредствованно отражают соотношения в фонологической системе указанных языков, так как алфавиты этих языков как-то соотносятся с их фонологическими системами, однако говорить о всеобщности преобладания дентальных (переднеязычных) во всех языках слишком рискованно. Позднейшие работы, в частности исследования чешского лингвиста Иржи Крамского[14], показали, что соотношения фонем по признаку места артикуляции различны в разных языках. Языки, обследованные Бурдоном, и по данным Крамского содержат больший процент переднеязычных, чем всех остальных согласных вместе взятых (т. е. больше 50% всех согласных). Однако в его таблице среди 23 обследованных языков имеются и такие, в которых переднеязычных меньше 50%. В языке арапахо (Северная Америка) переднеязычных 41,7%, а заднеязычных — 48,5%; в языке гола (Африка) переднеязычных всего 33,8%, встречаются и другие языки с пониженной долей переднеязычных.

Это показывает, с какой осторожностью следует относиться к подобным обобщениям. К сожалению, данные Крамского также недостаточны (для каждого языка выборка не превышает 3000 фонем), и он сам это сознает, однако для опровержения мнимого закона они вполне применимы.

Бурдон считал, что наблюденные им соотношения в частоте элементов для разных языков находят свое отражение в развитии человеческих эмоций и психики вообще от животного начала, связываемого им с лабиальной и велярной артикуляцией, к началу человеческому — артикуляция дентальная, альвеолярная и палатальная. Нет нужды останавливаться на опровержении этих концепций. Отметим, однако, что книга Бурдона для своего времени была полезной, так как обращала внимание на возможность квантитативной типологии языков.

---

[13] M. Weiss. Указ. соч., стр. 45: «...Verbleibt doch zweifellos die Bevorzugung den dentalen Laute in allen Sprachen. Das Gesetz von Bourdon über die Bevorzugung den dentalen kann sowohl auf der grossen Beweglichkeit der Zungenspitzen, als auch der relativ grösseren Zahl der Dentallaute im Verhältnis zu lauten anderer Artikulationsgebiete beruhen».

[14] J. Kramský. A quantitative typology of languages. «Language and Speech», 1959, v. 2, стр. 11.

После работы Бурдона психо-биологическая интерпретация частотности звуковых элементов стала опираться на новые данные, открывшиеся в результате быстрого развития в начале века экспериментальной и общей фонетики. Это было связано с исследованиями Суита, Свифта, Джоунза, ван Гиннекена и других фонетистов, с возникновением принципов Международной фонетической ассоциации, подчеркивавших необходимость единого подхода к описанию звуков всех языков. В этом плане показательна теория фонетической частотности голландского фонетиста И. ван Гиннекена [15]. Согласно этой теории, имеется определенное конечное число физиологических движений речевого аппарата — артикуляций, причем артикуляции некоторым прямым образом соответствуют антропологическим расовым признакам. Различия в системах фонем и их частотности в различных языках объясняются различным распределением (в результате расового смешения) артикуляторных признаков. Здесь ван Гиннекен оказался под определенным влиянием физической антропологии, внутри которой к 10—20-м годам методы математической статистики применялись достаточно широко; распространив эти методы на языкознание, он привлек и весь концептуальный аппарат тогдашней антропологии.

Убедительную критику концепций ван Гиннекена дал в «Основах фонологии» Н. С. Трубецкой. Он писал: «Эта теория создана не индуктивным путем, она исходит не столько от 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 ином языке имеет особенно высокую или низкую частотность, то строится предположение, что расовые признаки 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 северо-американских индейцев, то это еще далеко не свидетельствует о зависимости фонемной частоты от расовых приз-

---

[15] J. van G i n n e k e n. Die statistiek in taalwetenschap. «De Nieuwe Taalgids», Groningen, 1915, стр. 65—95; О н ж е. Benutzung der statistischen Methoden für die Sprachwissenschaft. «Indogermanisches Jahrbuch», Bd 10, стр. 43; О н ж е. Ras en Taal. «Verhandelingen der Koninglijken Akademie van Wetenschapen te Amsterdam», Aft. Letterkunde, 1935, № XXXVI; О н ж е. De Ontwikkelingsgescheidenis van de systemen der menschelijke Taalklanken. Amsterdam, 1932; О н ж е. De Oorzaken der taalveranderingen. Amsterdam, 1930; О н ж е. La biologie et la base d'articulation. «Journal de Psychologie», 1932, XXX, стр. 266—320.

наков, поскольку языки негров отличаются от языков северо-американских индейцев не только разной частотностью, 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 частоту 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 разных языков. Только тогда имело бы смысл обсуждать этот вопрос» [16].

Теория ван Гиннекена, таким образом, оперировала звуковыми элементами и их частотами не как лингвистическими, смыслообразующими элементами, а как физиологическими признаками. Она была последней попыткой найти непосредственное психобиологическое причинное обоснование различия в частотности звуковых элементов в языках.

* * *

Как известно, фонология, зародившаяся в Пражском лингвистическом кружке, пришла к постулату о том, что физически сходные звуковые элементы, будучи помещены в различные системы отношений с другими элементами, оказываются функционально, лингвистически различными.

Идеи, близкие идеям фонологии, начали зарождаться и в лингвистической статистике, которая в середине 20-х годов складывается в самостоятельную отрасль лингвистики. Окончательно сформулировал основные положения лингвистической статистики (в том числе и применительно к фонологии) Дж. К. Ципф, по праву считающийся основателем современной квантитативной лингвистики [17].

Понятийный аппарат своих исследований Дж. К. Ципф черпал из двух источников: современной ему психологии (бихевиоризм

---

[16] Н. С. Трубецкой. Основы фонологии. М., 1960, стр. 291.

[17] G. K. Zipf. Relative Frequency as a Determinant of Phonetic Change. «Harvard Studies in Classical Philology», № 40. Cambridge, Mass., 1929; Он же. Selected Studies of the Principle of Relative Frequency in Language. Cambridge, Mass., 1932; Он же. The Psycho-Biology of Language. Boston, 1935; Он же. Human Behaviour and the Principle of the Least Effort. N. Y., 1946; Он же. Statistical methods and dynamic philology. «Language», 1937, v. 13, стр. 60—70; Он же. Homogeneity and heterogeneity in language. «Psychological Records», Bloomington, 1938, № 2. Он же. Phonometry, phonology and dynamic phylology: an attempted synthesis. «American Speech», v. 13, N. Y. 1938; Он же. The Psychology of Language. P. L. Harriman (ed.). «Encyclopedia of Psychology». N. Y., 1946.

и гештальтпсихология) и классического сравнительно-исторического языкознания, как оно сформировалось к концу XIX — началу XX в. При этом Ципф стремился объединить эти два потока в единое лингвистическое учение — динамическую филологию. В этом и сильная и слабая стороны концепции Ципфа. Обе эти стороны связаны с особенностями научной личности Ципфа. Он был достаточно крупным ученым, чтобы создать свое направление в языкознании, однако он не был компаративистом (и лингвистом вообще) на уровне 30-х годов нашего века, его лингвистический кругозор был ограничен, поэтому «динамическая филология» как заменитель всего языкознания не удалась. В стремлении всюду находить Общие Принципы Ципф был до некоторой степени провинциален и, пожалуй, главное — он оперировал ограниченным кругом материалов. Все они содержатся в двух его первых монографиях.

С другой стороны, Ципф обладал поразительной интуицией и последовательностью, что позволило ему поставить ряд очень важных вопросов и в общих чертах наметить современную проблематику фонологической статистики.

Психологическая концепция Ципфа достаточно объемлюща. «Весь опыт — это реакция, структурно организованная в своем источнике. Любая реакция — есть выражение, как только она осознается, а любое выражение — это язык, как только нам удается расшифровать его. То, что мы обычно называем языком,— это весьма частная область поведения, чей код достаточно хорошо известен» [18]. Эта формулировка замечательна тем, что ставит в общем виде проблемы соотношения языка с другими системами коммуникации. Подобное рассмотрение языка как совокупности знаков (по терминологии Ципфа — актем), соотнесенных с определенным значением («гены значения»), в каком-то смысле сходно с постулатом о знаковой природе языка, высказанным де Соссюром в начале века, и как раз в 20-е годы подхваченным пражской школой. Ср. следующее высказывание Ципфа: «...Сознательная деятельность разума относится к бессознательной физиологической деятельности как предложения относятся к фонемам» [19]. Это противопоставление сознательного разумного и бессознательного физиологического может быть сопоставлено с дихотомией означаемого и означающего, плана содержания и плана выражения. Различие между подходом Ципфа и концепциями структурной лингвистики, складывавшимися в 20-е годы, состоит в том, что Ципф был слишком приближен к психологии, его не интересовали внутренние проблемы лингвистики, вернее, он видел их разрешение в проблемах внелингвистических.

---

[18] G. K. Z i p f. The Psycho-Biology of Language, стр. 309.
[19] Там же, стр. 302.

Помещая язык в общий контекст поведения, Ципф следующим образом формулирует свою основную гипотезу: «...на всех уровнях — внутриличностном и межличностном, материальном, физиологическом и ментальном предполагается состояние органического равновесия, которое мы стремимся поддерживать, анализируя и реорганизуя новый опыт в заранее установленные классы и конфигурации» [20]. Для языка гипотеза о стремлении системы к равновесию излагается следующим образом: «все речевые элементы или языковые структуры с неизбежностью подчиняются в своем поведении фундаментальному закону экономии, который состоит в стремлении поддерживать равновесие между формой и поведением» [21]. Под формой понимается структура, строение элемента, а под поведением — его встречаемость (occurence). В дальнейшем «закон экономии» будет сформулирован Ципфом как «принцип наименьшего усилия», характеризующий, по его мнению, все человеческое поведение. Этот ципфовский принцип часто толкуется как тенденция к выработке в процессе коммуникации наименее избыточного кода.

В настоящее время теоретико-информационные исследования показали, что код естественного языка обладает очень большой избыточностью, однако эта избыточность необходима для достижения большей помехоустойчивости и гибкости коммуникации. Таким образом, общий принцип экономии Ципфа не является столь всеобъемлющим и его не следует трактовать как автоматическую тенденцию к уменьшению избыточности в языке.

Лингвиста не должна смущать психологическая терминология Ципфа или его стремление объяснить все некоторыми общими принципами. В главном его концепция представляется нам чрезвычайно поучительной и во многом перспективной даже сейчас. Это — его стремление связать лингвистическую эволюцию с живыми речевыми процессами, чего до него не было. Попробуем изложить некоторые фонологические идеи Ципфа в том виде, как они изложены в его первой монографии «Относительная частота как определяющий фактор языковых изменений». Обычно при изложении идей Ципфа прибегают к его более поздним работам; нам представляется, что это не всегда удачно, так как в более поздних работах наблюдается усиленная тенденция к психологизации и чрезмерному обобщению, а в своей первой работе Ципф еще не испытал никакого влияния со стороны, его больше интересуют конкретные фонологические проблемы, чем общеязыковые спекуляции. Кроме того, интересно следить за ходом мысли исследователя, которая привела его к тем или иным выводам.

Свой принцип частотности Ципф выводит из наблюдений над поведением слов в связной речи: «Ученые уже давно отметили,

---

[20] G. K. Zipf. The Psycho-Biology of Language, стр. 303.
[21] Там же, стр. 19.

что если слово, ранее употреблявшееся редко, начинает неожиданно употребляться гораздо чаще, то его форма может с большой вероятностью упроститься, чтобы слово было легко произносить»[22].

Далее формулируется принцип частотности:

«Ударение, или степень выделенности любого слова, слога или звука обратно пропорциональны относительной частоте этого слова, слога или звука среди других соответствующих слов, слогов или звуков в потоке речи. С увеличением употребительности форма становится менее акцентированной или более легко произносимой»[23].

Разумеется, весьма уязвимым в этой формулировке является критерий легкости произнесения. Н. С. Трубецкой следующим образом анализирует этот критерий: «...со строго естественнонаучной 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 противопоставления по способу преодоления преграды. В еще меньшей мере может быть определена степень сложности в отношении противопоставлений по месту образования преграды. Ципф в качестве примера приводит противопоставление *m* — *n* и, исходя из того 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 небной 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

---

[22] G. K. Zipf. Relative Frequency as a Determinant of Phonetic Change, стр. 3.

[23] Там же, стр. 4.

отвергнута, по крайней мере в приведенной выше редакции» [24].

Аргументация Трубецкого в высшей степени убедительна. Однако стоит поближе присмотреться к положениям Ципфа, чтобы убедиться, что в его понимании (может быть, имплицитно) сложность и легкость произношения не носили того физиологически-артикулярного характера, на котором основывается вся критика Трубецкого. Думается, что Ципфа не интересовала мышечная легкость или сложность; его понимание этих понятий приближалось скорее к фонологическому. Приведем его высказывания по этому поводу:

«Рассматривая все множество согласных языка как определенное количество сочетаний фонетических звуков, мы сразу обнаруживаем, что одни более трудны для произношения, чем другие. Возьмем для примера дентальные: *dh* и более труднопроизносимо, и более слышно, чем *d*, поскольку оно содержит все то же, что и *d*, плюс добавочный элемент аспирации. Таким же образом звонкое *d* труднее для произношения, чем глухое *t*, оно также более выделено и слышно, поскольку *d* имеет как дентальное место образования, так и взрывность *t* плюс добавочный элемент звонкости. По подобным причинам придыхательное *th* и аффриката *ts* более выделены и более слышны, чем *t*. Но как насчет спиранта θ? Он кажется наименее выделенным из всех, так как имеет лишь межзубное место образования без добавочного элемента звонкости, или аспирации, или взрывности. Следует, однако, заметить, что отсутствие взрывности, звонкости и аспирации здесь компенсируется продолжительностью (θ, θθ, θθθ и т. д.) — качеством, почти полностью отсутствующим у других рассмотренных звуков.

Подобно дентальным ведут себя лабиальные и гуттуральные. Это можно, по-видимому, обобщить следующим образом: звонкие придыхательные являются фонетически и акустически более выделенными, чем звонкие; звонкие более выделены, чем глухие, аффрикаты более выделены, чем глухие и, возможно, гораздо более отличаются от глухих, чем звонкие. Спиранты могут быть выделены, а могут оставаться невыделенными в зависимости от их продолжительности. Можно в принципе утверждать, что каждый согласный состоит из определенного числа единиц фонетической трудности или акустической выделенности» [25].

То, как Ципф понимал природу и строение фонологических элементов, показывает, что здесь имеются в виду не конкретные звуки речи — фонетические звуки являются составными частями этих элементов,— а некоторые более абстрактные единицы. Собственно говоря, в своей книге «Психобиология языка» Ципф

---

[24] Н. С. Трубецкой. Указ. соч., стр. 292.
[25] G. K. Zipf. Relative Frequency as a Determinant of Phonetic Change, стр. 36.

вводит понятие фонемы. Как мы видим, уже в первой монографии он, хотя и не формулируя этого, имеет дело также с фонемами. В «Психобиологии языка» имеется понятие значимого акта поведения, основанное на гештальтпсихологических принципах. Это понятие вполне тождественно понятию различительного признака, возникшему в фонологии и позднее ставшему фундаментальным не только для лингвистики, но и для многих других наук [26].

Но ведь понятия добавочных элементов (звонкость, продолжительность, аспирация и т. п.), используемые еще в первой работе Ципфа, также не обозначают физиологических артикуляторно-мышечных явлений, а представляют собою абстрактные различительные признаки. Это видно из того, что фонологические элементы понимаются как пучки дискретных единиц — вполне аналогично тому, как сейчас фонема понимается как пучок различительных признаков (a bundle of distinctive features) [27]. Более того, очевидно, что Ципф понимал структуру фонологического элемента как наличие—отсутствие того или иного признака, т. е. дихотомически. Таким образом, критика Трубецкого относится скорее лишь к одной из возможных интерпретаций Ципфа (не самой вероятной). Сам же Ципф близко подходил к чисто фонологическим идеям, причем в их более современном виде («теория различительных признаков»). Во всяком случае, его теория не нуждается в физиологически-артикуляторных обоснованиях (непосредственно). Единицы, которые Ципф рассматривал, и их строение суть единицы фонологического уровня.

Вот как описывает Ципф действие принципа частотности, т. е., пользуясь его терминами, ослабление выделенности слова с увеличением его относительной частоты:

«Предположим, например, что в каком-то языке каждое произнесенное слово начинается с *d*. Что произойдет? Несомненно, это *d* перестанет быть характерной частью слова (разрядка моя.— *Д. С.*); говорящий, стремясь сократить усилия на произнесение, будет игнорировать этот звук (Trägheitsgesetz); слушающий, привыкнув к этому звуку, не будет настаивать на его четком произношении. В результате начальный звук *d* станет более слабым и перейдет в другой, более легко произносимый дентальный звук, а может быть, и исчезнет вовсе. Такое же ослабление *d* произойдет, если он будет занимать не исключительно инициальное, а, положим, исключительно медиальное или исключительно финальное положение. Нам кажется, что этот тезис неоспорим.

---

[26] Ср., в частности: Jean Viet. Les méthodes structuralistes dans les sciences sociales. Paris — den Haag, 1965.

[27] Ср.: R. Jakobson, M. Halle. Fundamentals of Language. 's-Gravenhage, 1956.

Но действительно ли необходимо, чтобы в потоке речи наше *d* занимало какое-то определенное место в слове (инициальное, медиальное или финальное), чтобы оно стало ослабевать? На самом деле *d* в таких обстоятельствах будет ослабевать, если оно будет содержаться просто в очень большой части слов языка. Пойдем, однако, далее и предположим, что половина или три четверти или другая большая доля употребляемых согласных занята *d*, которые могут стоять на любом месте в слове. Я имею здесь в виду не согласные слова в словаре, где каждое слово встречается лишь один раз, а согласные так, как они встречаются в потоке речи. Другими словами, предположим, что сто самых частых слов содержат *d* — вся наша работа имеет дело с языком в потоке живой речи. Итак, если в речи *d* будет составлять половину или три четверти всех согласных, то оно несомненно подвергнется ослаблению. Каков этот верхний порог, мы не знаем, но представляется весьма правдоподобным, что он существует. Поэтому если германское *d* содержалось в речи именно в такой пропорции, это прекрасная причина, по которой оно могло перейти в глухое *t*. Перейдя теперь к дентальному *t*, которое является на один шаг более слабым, чем *d*, мы обнаружим, что этот звук, будучи менее трудным для произнесения и менее заметным, должен, соответственно, встречаться в потоке речи гораздо чаще, чем *d*. Его более частое повторение не будет столь заметным, как повторение *d*, и этот звук не перестанет быть характерной частью слова. Если, однако, процент *t* превысит некую норму, то он должен будет, в свою очередь, подвергнуться ослаблению. И если бы удалось доказать, что в общегерманском процент *t* был выше нормы, это бы прекрасно объяснило ослабление *t* в ѳ. Аргументы, подобные вышеприведенным, можно было бы привести и относительно других звуков и показать, что у каждого звука должен быть верхний порог частоты, который он не может превысить, чтобы не оказаться ослабленным.

Однако наше рассуждение можно провести и в обратном порядке. Предположим, что в потоке речи встречается очень мало звуков *t*. Настолько мало, что они почти не употребляются. Тогда эти звуки станут отчетливой различительной частью слова, которую говорящий будет стремиться произносить отчетливо, а слушающий будет отделять от других звуков. Внутреннее ударение (inner accent, Hauptgestalt) слова будет сосредоточиваться вокруг *t*. Поэтому возможно, что говорящий, стремясь произносить его четко, будет бессознательно прибавлять к звуку дополнительный элемент аспирации или спирантизации и т. п. Итак, любое *ts* или *th*, возникшее из *t*, ясно показывает, что говорящий делает дополнительное усилие на *t*. Другими словами, имеется не только верхний порог частоты для, допустим, *t*, но и нижний порог, ниже которого *t* будет стремиться принять более отчетливую форму. Таким образом, могут возникнуть придыхательные или

аффрикаты, или, скорее, таким образом звук может принять более отчетливую форму» [28]

Основной аргумент идет, как мы видим, от частоты к структуре фонемы. Ципф далее проделывает целый ряд подсчетов с целью доказать свой тезис. Некоторые из этих данных представляются весьма убедительными и вносят, таким образом, новый элемент в тогдашнее сравнительное языкознание. Вот некоторые наиболее удачные примеры.

На своих подсчетах из латинского языка Ципф убедился в том, что лат. *m* имеет частоту 5,82%, а *n* — 6,47%, т. е. частота *m* гораздо более приближается к частоте *n*, чем во всех других обследованных им языках. Что это отражает действительное положение вещей, ясно из того, что в латинском языке элемент *m* входит в состав весьма употребительных именных и глагольных флексий. Из факта превышения *m* некоторого порога частоты делается вывод, что это было причиной его устранения в формах аблатива, а также в номинативе среднего рода и некоторых других формах.

Другой пример представляется нам еще более показательным, так как он дан в непосредственном наблюдении. По данным частотного словаря Кединга, которые Ципф перегруппировал так, чтобы они описывали уровень не букв, а звуков, в немецком языке частота *n* 10,4%, т. е. гораздо выше, чем в других языках. Вспомним, что в немецком языке *n* является основным формообразующим элементом как в имени, так и в глаголе. Соответственно во многих немецких диалектах, особенно северных, *n* в в грамматических окончаниях глагола, прилагательного и существительного элиминируется. Сравним аналогичную ситуацию в голландском языке, где этот процесс стал уже частью литературной нормы: kinderen ['kindərə]— 'дети' и т. п., а также ср. английский язык, где все *n* во всех флексиях были устранены еще в среднеанглийский период.

Таким образом, представление Ципфа о ходе звуковых изменений оправдывается по крайней мере в некоторых случаях. Отметим, однако, что эти случаи относятся к полному (или почти полному) устранению элемента. Что же касается изменения его фонологического состава, здесь ситуация далеко не так проста, как рисует Ципф. Приведем пример также из германских языков. В немецком и голландском частота *g* выше, чем в других языках. Причины этого ясны — и в том и в другом языке этот элемент входит в состав весьма употребительного страдательного причастия. При этом в голландском языке *g* имеет характер спиранта, а в литературном верхненемецком языке — характер взрывного. Спирантизация *g* в голландском могла бы быть объяснена вслед

[28] G. K. Zipf. Relative Frequency as a Determinant of Phonetic Change, стр. 38—40.

за Ципфом его ослаблением вследствие высокой частоты, однако в таком случае остается непонятным, почему это не произошло в немецком при относительно равных условиях.

Здесь мы подходим к основному пороку концепции Ципфа. Дело в том, что, оставаясь на позициях старого фонетизма, он рассматривал физически сходные элементы — звуки — как одинаковые для всех языков. Для Ципфа еще не существовала система отношений фонем, хотя, как мы пытались показать, сами фонемы он рассматривал как абстрактные элементы. Каждая фонема существовала отдельно как элемент, не связанный с другими в системе данного языка. Все фонемы были одинаковы для всех языков (*t* в английском приравнивалось *t* в немецком, *t* в русском и т. п.). Поэтому принцип частоты формулировался не для данной системы, а для глухих (или звонких) вообще, абстрагируясь от того, в какие соотношения с другими элементами в системе они входят.

Понимание природы фонетических признаков, как видно из приведенной цитаты (см. стр. 30), было у Ципфа также весьма односторонним. Хотя «добавочные элементы» и понимались им как абстрактные признаки, ему мешала трактовка наличия признака как усиления, а его отсутствия — как ослабления звука (или его слышимости). Даже если не понимать признаки в мышечно-артикуляторном плане, то все равно введение «силы—слабости» в качестве идентифицирующего критерия неудачно, в частности, потому, что это вносит путаницу во внутренний фонологический признак «сильный—слабый» (или «напряженный—ненапряженный» — tense—lax). Например, явно неудачна трактовка противопоставления «звонкость—глухость» как аналогии противопоставлению «сильный—слабый» при том, что данные прежде всего германских языков (не говоря об уральских, которые Ципф не исследовал) противоречат этому. В частности, для немецкого языка *t* никак не может считаться слабым, а *d* — сильным (аналогично и в других парах).

Следует отметить, что сам Ципф чувствовал слабость введенного им понятия «ослабления» или «усиления». Разбирая пример с озвончением интервокального *p* при переходе нар. лат. ripa в провanс. riba, Ципф замечает, что *p* перешло в *b* не из-за малой частотности *p* в этой позиции, а наоборот, из-за его большой частотности. Это противоречие принципу частотности («слабый» звук *p* перешел в более «сильный» *b* под влиянием увеличения частоты) объясняется тем, что здесь происходит не усиление (как при аналогичном переходе для других пар «глухой—звонкий»), а ослабление: «в интервокальном положении *b* произнести легче, чем *p*». Гештальт слова (по объяснению Ципфа) требует, чтобы *p* перешло в *b*. Знаменательно то, что автор вынужден возвратиться здесь к слову как к основному локусу фонетических процессов; вспомним, что свое изложение принципа частотности Ципф начинал с установле-

ния релевантной позиции различения в слове (начало слова, конец, середина). Последующий отказ от использования понятия значимой позиции во многом ослабил аргументацию Ципфа. Отметим, что в разобранных нами примерах элиминации латинского *m* или германского *n* эта элиминация происходит в определенных морфонологически релевантных позициях в слове; иными словами, увеличение встречаемости фонетического элемента обычно связано с увеличением встречаемости определенного грамматического элемента (а не лексемы), так как частота этих элементов превышает частоту лексем (формант прошедшего времени в русском языке, например, встречается чаще, чем каждый глагол в отдельности). Грамматический элемент, со своей стороны, связан с определенной позицией в слове (ср. суффигирование грамматических элементов во многих языках, в частности в русском, в противоположность префигированию в австроазиатских языках).

Конечно, теория звуковых изменений Ципфа не объясняет всего комплекса факторов, влияющих на план выражения. Эта теория описывает одну сторону языковой деятельности. В общем виде она соотносится с идеями представителей пражской фонологии и в первую очередь Р. О. Якобсона о том, что звуковые изменения начинаются в так называемом беглом стиле речи (аллегростиль)[29]. Однако Ципф совершенно не учитывает, что изменения на фонологическом уровне часто происходят в силу чисто системных факторов. Отдельные фонологические устремления Ципфа оказались недостаточными, чтобы быть связанными в цельную фонологическую теорию.

До сих пор мы прослеживали одно рассуждение Ципфа — о влиянии увеличения или уменьшения встречаемости на форму звука. Из общеметодологического представления о существовании равновесия в системе Ципф делает вывод о том, что, соответственно, более слабые звуки должны в тексте встречаться чаще, чем более сильные, т. е. что неаспираты должны встречаться чаще, чем аспираты, а глухие — чаще, чем звонкие. Для того, чтобы доказать этот тезис, Ципф привлек частотные данные по 10 языкам: французскому, русскому, чешскому, болгарскому, немецкому, испанскому, итальянскому, шведскому, венгерскому и английскому. Как уже отмечалось выше, данные по некоторым языкам (русский, итальянский) заимствованы Ципфом из книги Бурдона, где они основаны на алфавитных текстах очень малого объема (3 000—4 000 букв); равным образом неудовлетворительны данные по болгарскому и испанскому языкам. Достоверные данные относятся лишь к английскому и, до некоторой степени, немецкому языкам. Однако алфавиты всех указанных языков в какой-то мере отражают их фонологические системы, а поскольку цели Ципфа были

---

[29] Ср., в частности, статью: Г. А. Баринова. О произношении [ж'] и [ш']. «Развитие фонетики современного русского языка». М., 1966, стр. 54.

ограничены, он сумел доказать свой тезис, а именно, что в исследованном им материале глухие согласные встречаются чаще звонких.

Именно в таком виде принцип частотности Ципфа стал известен в лингвистике; положение об ослаблении формы звука в результате возрастания его употребительности, бывшее первичным, отошло на задний план. Следует отметить, что это второе (если угодно, обратное) понимание принципа Ципфа, которое для самого автора было лишь иллюстрацией первого, с нашей точки зрения, не соотносится с первоначальным тезисом зеркально, как считал Ципф. Дело в том, что эмпирический факт превосходства частоты некоторых групп фонем над другими подтверждается на самых различных языках, однако, как мы уже отмечали, принцип частотности в его первоначальной редакции подтверждается далеко не всегда, и не всякие фонетические изменения можно им объяснить. Поэтому представляется, что оба толкования принципа частотности надо отделить друг от друга и рассматривать в качестве исходного эмпирический факт, установленный Ципфом. Так интерпретирует данные Ципфа Н. С. Трубецкой, который понял, что подход Ципфа хотя и не является в основе фонологическим, но может быть истолкован с точки зрения фонологии (думается, что подобное истолкование возможно потому, что у самого Ципфа есть элементы фонологического подхода). Н. С. Трубецкой писал:

«В фонологической редакции эта теория могла бы звучать примерно так: «из двух членов привативной оппозиции немаркированный член 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 не нейтрализуемые 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 — *pbfvtdszmnr* фактически встречаются гораздо чаще, чем соответствующие им мягкие: *p'b'f'v't'd's'z'm'n'r'* (в отношении примерно 2 : 1). Но для 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, сибилянтами, палатализованными лабиальными) [30]. Корреляция звонкости в русском языке

[30] Фактические данные Н. С. Трубецкого о соотношении *l*—*l'* в русском языке оказались неверными, что иллюстрирует нашу мысль о необходимости надежных исходных подсчетов. Мы в настоящей работе использовали материал Генри Кучеры, опубликованный в его докладе на

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 шумных. Иными словами, фонема *v* оказывает на другие шумные совсем не такое 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 членах рассматриваемой корреляции звонкая фонема встречается примерно в три раза реже, чем глухая [31]

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: здесь придыхательные смычные встречаются, как правило, чаще, чем соответствующие им непридыхательные.., и только у глубоковелярных отношение оказывается обратным.., следует, однако, заметить, что в противоположность всем прочим оппозициям по интенсивности оппозиция q^h: Q в послеударных 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

V Международном съезде славистов в Софии (выборка 100 000 фонем). По данным Кучеры, это соотношение равно 0,0266—0,0208 (т. е. почти в точности соответствует ситуации в польском языке: $\mu - l = 0{,}027{:}0{,}021$). Как видим, здесь на самом деле *l* немного превышает *l'* по частоте, однако это ничуть не нарушает аргументации Н. С. Трубецкого об особом положении пары *l—l'* в отношении частоты среди других пар «непалатализованный — палатализованный». В то время как в других парах частота твердого превышает частоту мягкого почти в два раза, здесь эти величины очень близки, что вполне достаточно для доказательства отличия характера нейтрализации *l—l'* от остальных пар.

[31] На самом деле ($v + v'$) в четыре раза чаще, чем ($f + f'$). Что же касается того, что остальные глухие примерно в три раза чаще, чем звонкие, то это не совсем точно. Для некоторых пар это превосходжение меньше.

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 *а*), причем выбор представителя архифонемы обусловлен внешними обстоятельствами, в силу чего немаркированный характер какого-либо члена этой оппозиции не может быть установлен объективно. В общем же глухие шумные встречаются чаще соответствующих звонких.., но в каждом отдельном случае отношение оказывается разным» [32].

В этой цитате Трубецкой формулирует основную задачу фонологической статистики: найти такую фонологическую модель, которая бы адекватно описывала эмпирические наблюдения Ципфа, что некоторые звуки более часты в языке, чем их корреляты. Таким образом, работа Ципфа поставила проблему, которая до настоящего времени остается в центре фонологической статистики. В нашей работе мы попытаемся дать свое понимание правила Ципфа.

* * *

Выход в свет первых работ Ципфа совпал с началом деятельности Пражского лингвистического кружка. Общее отношение пражской фонологии к проблематике лингвистической статистики, и в частности к работам Ципфа, хорошо суммировано Н. С. Трубецким в его книге «Основы фонологии», обширные цитаты из которой приведены выше. Трубецкой совместно с В. Матезиусом [33], а также Б. Трнкой [34] сформулировал основные положения фонологической статистики в свете требований пражской школы. Эти положения сводятся к следующему.

1) При описании фонологической системы языка выявляется эмпирический факт, заключающийся в том, что одни оппозиции распространяются на подавляющую часть фонемного инвентаря, а другие — лишь на ограниченную его часть, причем это ограничение действия оппозиции может коррелировать с ограничением частоты употребления членов, охватываемых этой оппозицией. Например, корреляция палатализованности регулярным образом охватывает систему русского консонантизма, а если взять в качестве противоположного примера кетский язык, то в нем существует корреляция инъективности, которая, однако, выступает лишь в определенной (финальной) позиции и охватывает не

[32] Н. С. Трубецкой. Указ. соч., стр. 292—293.

[33] V. Mathesius. La structure phonologique du lexique du tchèque moderne.— TCLP, I. Praha, 1929, стр. 67—85; Он же. Zum Problem der Belastungs- und Kombinationsfähigkeit der Phoneme.— TCLP, IV. Praha, 1931.

[34] B. Trnka. A phonological analysis of present-day standard English. «Práce z vědeckých ústavů», XXXVII. Prague, 1935, стр. 45—175.

все согласные. С этими и подобными фактами связано понятие функциональной нагрузки противопоставления (и фонемы), введенное пражской школой. Сюда же примыкает и фонологическая интерпретация правила Ципфа.

В. Матезиус в своей работе о функциональной нагрузке и комбинаторных возможностях фонемы писал: «Для фонологической характеристики языка недостаточно назвать инвентарь фонем и фонологических признаков; следует при этом исследовать интенсивность, с которой употребляются отдельные фонологические единицы данного языка. Таким образом, фонология не ограничивается качественным анализом, но привлекает и количественный»[35].

В дальнейшем на выяснении функциональной нагрузки оппозиций в разных языках будут строиться попытки фонологической квантитативной типологии.

2) Уже из преимущественного интереса пражской фонологии к функциональной нагрузке элементов видна разница в подходе к фонологической статистике у этой школы и Ципфа. Для Ципфа фонология была динамическим процессом, манифестировавшимся в потоке речи. Равновесие элементов, устанавливаемое в процессе, было чем-то, к чему надлежало прийти, что следовало установить. Отсюда — упор, делаемый Ципфом на исследование текста. Частота — отражение и доказательство равновесия или его нарушения.

Для пражской фонологической школы основным понятием является понятие системы фонем как единого структурного целого. Поэтому для этой школы, в отличие от Ципфа, система фонем существует до частотного анализа, который призван лишь уточнить описание. Для Ципфа частота есть объяснение, интерпретация определенных фактов, для пражцев сама частота подлежит объяснению. Поэтому возникает стремление лингвистически осмыслить данные о частоте фонем.

Данные о функциональной нагрузке фонем и противопоставлений дают характеристику собственно системе; сразу же возникает вопрос, откуда следует черпать эти данные, чтобы они характеризовали именно инвариант.

И здесь пражская школа предлагает новый подход, последовательно отражающий ее стремление представить фонологический уровень в виде строгой системы. Тексту с его изменчивостью и индивидуальным характером предпочитается словарь языка. Предполагается, что если подсчитать частотность фонологических элементов по словарю, то эти данные будут характеризовать некоторый стабильный, постоянный инвариант. Таким образом, лексическая статистика словаря становится одной из отличительных черт пражской фонологии.

3) Однако функциональная нагрузка фонем в словаре является явно недостаточной и односторонней характеристикой системы.

[35] V. Mathesius. Zum Problem..., стр. 148.

Поэтому данные пункта 2 предлагается дополнить данными о частоте в тексте. Эти данные должны бросить новый дополнительный свет на словарную статистику, показать, насколько совпадает или не совпадает функциональная нагрузка элементов системы в языке и речи.

4) И, наконец, что очень важно, вводится понятие теоретической частотности (в особенности у Трубецкого). Под теоретической частотностью Н. С. Трубецкой понимает такую величину, которая выводится из теоретических возможностей встречаемости данной фонемы (с учетом правил нейтрализации и сочетаемости). Точных способов определения теоретической частотности не предлагается, указывается (на ряде примеров), что приблизительно величину теоретической частоты можно определить для минимальной пары фонем, зная среднее число слогов в слове, а также позиции нейтрализации противопоставления. Получаем, что немаркированный член противопоставления должен встречаться примерно в три раза чаще, чем маркированный. Далее выясняются реальные частоты фонем в текстах и определяется степень отклонения действительного соотношения от теоретического. Таким образом, фактическая частотность фонем в тексте оказывается соотнесенной с параметрами, представляющими теоретическую вероятность в тексте, а также в словаре. Именно изучение расхождений между этими видами данных позволяет интерпретировать результаты.

Позднее будут предложены и другие методы определения теоретической частоты (в частности, в работах П. Гиро, см. ниже). Здесь отметим, что стремление соотнести фактические данные с некоторыми теоретическими данными по своей тенденции совпадает с теоретико-статистическим разделением экспериментальной частоты и теоретической вероятности. Правда, различие между математической статистикой и пражской фонологией заключается в том, что в пражской фоностатистике теоретические характеристики выводятся не из статистической функции распределения, а из внестатистических соображений. Впрочем, и в самой статистике уже начинают приходить к пониманию необходимости и теоретической значимости сравнения экспериментальных данных с некоторыми внестатистическими (или же эмпирически установленными) параметрами. Ср. высказывания Дж. Айера по этому поводу: «...действительно значимым является не отклонение от априорных частот, а отклонение от частот, которые были установлены эмпирически» [36].

Дальнейшая судьба фоностатистических идей пражской школы показывает, что развивались именно те идеи, которые были связаны с существом ее фонологической теории. Статистическое обоснование квантитативных исследований в русле этого направления не развивалось совсем. До второй мировой войны работы

[36] J. A y e r. Chance. «Scientific American», 1965, № 11, стр. 51.

в области фонологической статистики и квантитативной типологии были предприняты такими чешскими учеными, как Й. Вахек, И. Крамский и др.; после войны эти же ученые (в особенности И. Крамский) продолжили свои изыскания в области квантитативной типологии. Следует отметить, что идеи пражской школы о статистическом исследовании фонологической структуры словаря были значительно продвинуты и развиты немецкими фонологами Паулем Менцератом и Вильгельмом Майер-Эпплером.

Уже после войны, в 1949 г., вышла статья Б. Трнки [37], в которой дается суммарное изложение проблематики фонологической статистики в духе пражской школы. Эта статья подытоживает достигнутое в фонологической статистике еще перед войной, так что ее можно считать отражением состояния исследований в период пражского структурализма.

Согласно Трнке прежде всего необходимо определить частоту фонем и фонемных оппозиций как в словаре, так и в тексте. В первом случае необходимо определить частоту встречаемости отдельных фонем и оппозиций, которыми располагает язык для создания всего запаса слов и форм. Подобное исследование является необходимым дополнением к любому качественному анализу языка. Именно степень использования фонемных элементов и частота их встречаемости придают фонемам и фонемным противопоставлениям квантитативную значимость, без которой любое фонемное описание языка будет неполным. Трнка считал, что исследование степени употребительности (продуктивности) должно охватывать все фонологические признаки, которые имеются в системе.

Далее предлагается способ представления фонологической структуры словаря. Слова рассматриваются как цепочки фонемных элементов. Структура этих цепочек определяется следующим образом. Все слова разбиваются на несколько групп, в каждую группу попадают слова, состоящие из одинакового количества слогов, т. е. в одну группу объединяются все односложные, в другую — все двусложные слова и т. п. Далее каждая группа слов с равным количеством слогов подразделяется согласно числу фонем, из которых состоят слоги, и согласно сочетаемости фонем. Таким образом, становится сравнительно легко указать продуктивность всех фонологических элементов языка в разных позициях внутри слова как для индивидуальных типов слов, так и для суммы этих типов. Образуется своего рода структура фонологических отношений в словаре, существенно уточняющая прежние положения пражской школы о позициях релевантности, нейтрализации и т. п. Возникает возможность структурной типологии словаря на фонологическом уровне. Подобного рода типология

---

[37] B. T r n k a. K výstavbě fonologické statistiky. «Slóvo a slóvesnost», 1949, № 11, стр. 59—64.

будет разобрана немного ниже на примере работ Крамского и Менцерата — Майер-Эпплера.

Трнка считает, что статистика словаря должна быть обязательно дополнена статистикой высказывания (речи), базирующейся на исследовании связных текстов. Важно установить предел относительного постоянства частоты, который различен для различных фонемных элементов. Существование такого предела Трнка объясняет, в частности, тем, что в более обширных текстах слова распадаются на две группы: слова, повторяющиеся часто, и слова, встретившиеся лишь однократно или всего несколько раз. Чем длиннее исследуемые тексты, тем более явственной становится грань между этими двумя группами слов. Трнка приводит пример из частотного словаря немецкого языка Кединга: 66 наиболее частых слов заполняют почти половину (49,62%) словарного материала текстов в 11 млн. слов, 320 слов составляют 72,25% и 1000 слов — 87%. Делается вывод о том, что в чешском языке, более синтетическом, чем немецкий, предел постоянства частоты фонем лежит выше, чем в немецком, так как идентичные словоформы повторяются реже.

Другим фактором, играющим большую роль в фонемной статистике, является стиль и содержание высказывания. Трнка считает, что, очевидно, имеется взаимоотношение между стилем и повторяемостью фонологических элементов.

В этой связи упоминается о том, что более частая встречаемость фонемы [ə] в английском языке характерна для разговорного стиля речи. Немецкий язык — литературный — характеризуется большей частотой консонантных пучков на стыках слов (внутри сложных слов), чем разговорный и обиходный стили, содержащие гораздо меньше сложных слов.

Трнка отмечает в своей статье и факты, положенные в основу правила Ципфа, позднее интерпретированного Трубецким: немаркированные фонемы обычно оказываются более частотными, чем маркированные. Приводятся соответствующие примеры о частотности звонких и глухих в чешском, английском и русском языках, палатализованных и непалатализованных в русском языке и пр. Отмечается, что наблюдаются индивидуальные отклонения от этого правила, которые, согласно Трнке, следует объяснять специальными причинами: например, большая частота звонкого [ð] в английском несомненно связана с тем, что оно встречается в часто употребляющихся местоимениях и словах местоименного происхождения (this, that, then, there, the и т. п.), где ранее th было глухим. В словаре, однако, английское глухое [θ] чаще звонкого [ð].

В заключение своей работы Трнка замечает, что статистика уточняет или вскрывает проблемы качественного характера, особенно в тех случаях, когда исследуемая действительность не поддается непосредственному качественному анализу вследст-

вие своей чрезмерной сложности или гетерогенности. Эвристическая ценность статистики состоит, по мнению исследователя, в ее способности раскрывать различие между ожидаемыми и реальными числовыми характеристиками, что должно, как правило, приводить к пересмотру качественной интерпретации.

Следует отметить, что в своих основных постулатах работа Трнки обозначила некоторые направления будущего развития фонологической статистики, однако ее во многом слишком общий характер уже не удовлетворял состоянию лингвистики в 1949 г. На очереди стояли более детальные исследования, которые бы развивали не только общую концептуальную схему, но и конкретные приемы статистического анализа. В частности, у Б. Трнки, как и вообще у пражцев, недостаточно подробно обосновывается утверждение о том, что статистика словаря должна дать картину употребительности фонологических средств, которыми располагает язык для создания всего запаса слов и форм. Понятие «словарь» выступает как нечто неопределяемое и не учитывается то обстоятельство, что в современных литературных языках словарь представляет собой сложную совокупность многих подсистем. Поэтому представление о том, что инвентарь всех слов и форм не будет содержать повторяющихся элементов, оказывается неверным, а ведь именно на этом представлении и основывается сама идея словарной статистики — стремление выяснить встречаемости элементов в системе, где каждое комплексное образование (слово, лексема) представлено лишь однократно, установить инвентарь возможностей, не осложненных динамическими факторами развертывания речевого потока.

В частности, повторяемость в словаре связана с производными словами (например, всевозможные диминутивы, экспрессивные формы и пр.), словоформами одной парадигмы и т. п. Включать их в словарь для статистики или нет? Обычная лексикографическая практика дает отрицательный ответ на этот вопрос. Однако те решения, которые предлагают имеющиеся словари, для целей фонологической статистики являются совершенно неудовлетворительными. Например, для русского языка фоностатистика словаря даст несообразно большие значения для частоты фонемы [т'] вследствие того, что словарной формой глагола является инфинитив. Это отражает вовсе не фонологическую структуру русского словарного состава, а тот довольно случайный факт, что в лексикографической практике инфинитив избран репрезентантом глагола. С другой стороны, исключение словоформ из словарей может привести к тому, что определенные морфологические значимые позиции вовсе не будут учитываться при статистике. Если же все производные слова и словоформы учитывать при статистике словаря, то возникает проблема повторяемости, о которой было упомянуто выше. Таким образом, представляется очевидным, что для целей фонологической

статистики словарь не может быть представлен в том виде, в каком он существует в лексикографической традиции.

По-видимому, такой словарь должен представлять собой нечто вроде словаря лексических, словообразовательных и грамматических морфем с указанием соответствующих морфологических правил соединения морфем между собой. Однако до сих пор в реальной фоностатистике словаря все вышеприведенные соображения не учитывались, что повлияло на результаты подсчетов.

Вспомним, что о различных трудностях словарной статистики писал еще Н. С. Трубецкой:

«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»[38].

Относительно статистики связных текстов в самой статье Трнки ставится вопрос о возможном влиянии подбора материала на подсчеты. Таким образом, сама фонологическая статистика в ее наиболее «нестатистическом» представлении начинает приходить к мысли о том, что необходимы более специальные методы подбора материала и его оценки для того, чтобы получить параметры, характеризующие весь язык в целом, а не только его произвольную часть.

* * *

Дальнейшее развитие более специализированной линии фонологической статистики связано уже не с пражской школой фонологии, которая в первоначальном виде закончила свою деятельность где-то около 1949 г., а с появлением работ математиков, заинтересовавшихся лингвистикой как полем приложения статистических методов.

Мы вернемся к изложению некоторых работ, связанных с идеями пражской фонологии, там, где будем говорить о применении фоностатистики к типологии. Сейчас же перейдем к работам более специального характера, которые стали появляться в 40-е годы.

---

[38] Н. С. Трубецкой. Указ. соч., стр. 299.

Здесь необходимо прежде всего вспомнить о замечательной работе русского математика А. А. Маркова [39], который еще в начале века впервые применил статистические методы к лингвистическому материалу, проиллюстрировав открытые им «простые однородные цепи» на примере чередования гласных в тексте «Евгения Онегина».

Следует вообще указать, что введение в обиход статистических исследований лингвистического материала значительно обогатило проблематику самой статистики и смежных математических областей. Открытие «цепей Маркова» ввело в теорию вероятностных цепей дотоле неизвестный раздел. Другим вторжением лингвистики в теорию коммуникации был так называемый закон Ципфа. Речь идет не о правиле Ципфа в том виде, как оно было сформулировано применительно к фонологическому уровню, а о гораздо более общем законе зависимости частоты слова в тексте от его ранга (т. е. номера в списке по убывающей частоте), который принял вид $p_r = k_r^{-1}$, где p_r частота слова с рангом r, а k — константа, равная частоте слова с рангом 1. Мы не будем обсуждать здесь проблематику закона Ципфа в его традиционной формулировке для словаря, так как закон Ципфа был подвергнут исчерпывающему рассмотрению в работе Р. М. Фрумкиной «Статистические методы изучения лексики» (М., 1964), где ему посвящена целая глава. Мы присоединяемся ко всем выводам Р. М. Фрумкиной. Наши собственные соображения на этот счет можно найти в специальной статье [40].

Отметим лишь, что интерес многих математиков и лингвистов к закону Ципфа, связанный с развитием теории информации и проникновением математических методов в языкознание, а также с более специальными вопросами теории кодирования, способствовал «математизации» лингвистической статистики, привлек внимание ученых к другим аспектам количественной стороны речевой деятельности.

Уже к началу 40-х годов сложилась такая ситуация, что количество работ, не просто описывающих язык количественными методами, но и оперирующих применительно к языку такими понятиями математической статистики как распределение, случайная величина, вероятность, параметры распределения [41], было по крайней мере достаточным, чтобы предпринять первую попытку синтеза. Таким синтезом (а во многом и первым шагом) явилась замечательная книга одного из крупнейших английских теоре-

---

[39] А. А. Марков. Опыт статистического исследования текста романа «Евгений Онегин». «Изв. Российской имп. Академии наук», серия 6, т. 7. СПб., 1913.

[40] Д. М. Сегал. Некоторые уточнения вероятностей модели Ципфа. «Машинный перевод и прикладная лингвистика», № 5. М., 1961.

[41] См., в частности: C. B. Williams. A Note on the statistical analysis of sentence-length as a criterion of literary style. «Biometrika», 1940, v. 31, стр. 356.

тиков статистики Дж. Юла «Статистическое изучение литературного словаря» [42], которая во многих отношениях до сих пор остается непревзойденным образцом статистического анализа стиля. Хотя книга Юла совершенно не касается проблемы фонологии, однако ее значение и для развития фонологической статистики трудно переоценить. Здесь впервые была по существу создана концепция вероятностей модели речевой деятельности; впервые была поставлена проблема текста (в абстрактном смысле) как статистической совокупности; впервые было показано, как строются лингвистические распределения, и т. п.

После второй мировой войны повысился интерес математиков к применению математико-статистических методов для описания языковых явлений.

Если лингвисты пражской школы рассматривали лингвистическую статистику как способ изучения функционирования системы языка и стремились найти лингвистическое обоснование наблюденных количественных данных, то в работах математиков делается упор как раз на то, чего не было у пражцев,— на правила статистической работы с языком. При этом лингвистическая интерпретация зачастую не интересует исследователя. Следует сказать, что математики отнеслись более скептически, чем некоторые лингвисты, к возможностям применения статистики в языке: «Вообще говоря, не следует применять статистику в слишком многих областях лингвистики» [43]. Подобная сдержанность вызвана, во-первых, справедливым опасением, что из количественных данных будут сделаны поспешные и неправомерные выводы. Именно поэтому во многих статьях подчеркивается, что вопрос выборки является основным при такого рода исследованиях. С другой стороны, некоторый скептицизм по отношению к статистике объясняется пониманием того, что в то время как математическая статистика имеет дело со случайными процессами, языковая деятельность таковым процессом заведомо не является. Однако в начальных публикациях по введению в лингвистическую статистику никто не обратил внимания на важность установления того, что же в языке все-таки подчиняется правилам случая.

Среди первых методических работ необходимо упомянуть две, вышедшие в конце 40-х — начале 50-х годов. Это — статьи Д. У. Рида «Статистический подход к количественному лингвистическому анализу» [44] и А. С. Росса «Вероятностные проблемы филологии» [45]. Д. Рид обсуждает два вопроса: а) как надо собирать обширный

---

[42] G. U. J u l e. The Statistical study of Literary Vocabulary. Cambridge, 1944.
[43] E. B. N e w m a n. Statistical Methods in Phonetics. «Manual of Phonetics», ed. L. Kaiser. Amsterdam, 1957.
[44] D. W. R e e d. A statistical approach to quantitative linguistic analysis. «Word», 1949, № 5, стр. 235—247.
[45] A. S. R o s s. Philological probability problems. «Journal of the Royal Statistical Society», ser. B. 12, 1950, стр. 19—50.

материал, чтобы произвести верный анализ частоты лингвистических форм, и б) когда количественные различия лингвистического материала можно считать значимыми. Таким образом, проблема переводится из плоскости «толкования цифр» в плоскость правильного сбора цифр и установления их значимости.

Совершенно очевидно, что здесь применимы те же статистические приемы, что и в других приложениях статистики. Требуется оценить разброс экспериментальных частот. Д. Рид вводит в практику лингвостатистических исследований некоторые элементарные приемы статистического анализа: предлагается определять среднее уклонение частоты языковых объектов. Рид впервые вводит также понятие значимости, широко применяемое в статистике: выдвигается гипотеза о том, что частоты некоторых объектов должны быть сходными (т. е. что им должна соответствовать одинаковая теоретическая вероятность), а если экспериментальные данные показывают расхождения, то они объясняются случаем. Далее определяется экспериментальное расхождение среднего уклонения частот двух фонем, которое сравнивается с теоретической величиной, и если отклонение экспериментальной величины от теоретической настолько велико, что его нельзя объяснить случаем, то первоначальная гипотеза отвергается и делается вывод, что различие между частотами обеих фонем значимо.

Следует отметить, что метод Рида, конечно, слишком элементарен. Он позволяет сравнивать лишь две величины, а не все распределение. Кроме того, и так ясно, что две различные фонемы отличаются одна от другой. Существенно установить нечто о характере отличия одной системы фонем от другой, этого же метод Рида не дает.

Д. Рид замечает, что используемые им статистические приемы разработаны эмпирически для таких чисто случайных процессов, как подбрасывание монеты или костей, в то время как в языке на встречаемость событий оказывают влияние помимо случайных и другие факторы. Статистическая методика может указать на то, что в данном случае на величину частот в выборке воздействовал некий фактор помимо случайного, но она не может указать, каков этот фактор.

Статья Рида не содержала каких-либо конкретных соображений о фонологической статистике, тем не менее эта работа имела важное значение, так как она впервые обратила внимание лингвистов на то, что недостаточно подсчитать частоту какого-либо элемента в небольшом тексте и утверждать, что этот параметр характерен для всего языка.

В статье А. Росса также содержатся общеметодические положения и излагаются некоторые приемы статистического анализа. И здесь содержатся предостережения по адресу возможных злоупотреблений статистикой. В основном работа посвящена глоттохронологии.

К указанным двум работам примыкает и цитированная выше статья Э. Ньюмена из «Справочника по фонологии». Эту статью отличает выгодная простота и ясность изложения, хотя иногда эта простота ведет к игнорированию серьезных проблем.

Нам хочется отметить работу Ньюмена, так как в ней в явном виде поднимается вопрос об однородности исследуемого материала.

* * *

Начиная с 50-х годов лингвистическая статистика конституируется в отдельную дисциплину со своей научной традицией и проблематикой.

Во Введении мы коснулись некоторых теоретических разногласий между противниками применения статистических методов в лингвистике и их сторонниками. Теперь настал момент рассмотреть некоторые практические результаты применения статистических методов в фонологии. Мы затронем не все напечатанные работы (их слишком много), а те, которые внесли особый вклад в развитие лингвистической статистики.

Традиционной областью приложения статистических методов стала фонологическая типология языков. Следует отметить, что различные исследователи по-разному определяют типологию — от простого сопоставления фактов различных языков до сравнения глубинных моделей, отражающих взаимосвязь языковых уровней.

Наиболее простой и распространенный способ применения количественных методов в фонологической типологии — это сравнение разных языков по количеству элементов, входящих в фонемный инвентарь, а также сравнение употребительности этих элементов в текстах для различных языков.

Подсчет соотношения различных элементов в фонемном инвентаре был впервые, как мы упоминали, выполнен еще Б. Бурдоном[46]. Он условно расклассифицировал языки согласно относительной частоте гласных и согласных фонем в инвентаре исследуемых языков на три типа: консонантные, вокалические и смешанные.

И. Крамский в 1948 г. выступает с существенным дополнением[47]. Он показывает, что по-настоящему можно говорить о вокалических и консонантных типах языков, лишь имея в виду употребление фонем в тексте. Крамский поэтому предлагает ввести новую величину для вокалического коэффициента $v = p_i/p_t$, где p_i — процент согласных в системе, а p_t — процент согласных в тексте. Согласно этому показателю языки распределяются по вокалическому и консонантному типам следующим образом:

---

[46] B. Bourdon. Указ. соч., стр. 64.

[47] J. Kramský. Fonologické využití samohláskových fonémat. «Linguistica Slovaca», IV—VI. Bratislava, 1946—1948, стр. 39 и сл.

немецкий	0,85	чешский	1,15
английский	0,91	персидский	1,36
словенский	1,04	испанский	1,46
англо-саксонский	1,12	итальянский	1,58

Немецкий оказался наиболее, а итальянский — наименее консонантным.

К сожалению, к этой работе, так же как и к последующей, которую мы уже упоминали [48], следует предъявить серьезный упрек в недостаточности материала. Если цифры процента гласных в инвентаре не требуют статистической проверки, то соответствующие величины для текста нуждаются хотя бы в грубой верификации. Этого не сделано, поэтому работу Крамского можно воспринимать скорее как заявку, чем выполненное исследование.

Во второй из упомянутых работ автор исследует небольшие (размером до 3867 фонем) тексты на 23 языках и сравнивает относительные частоты отдельных классов фонем, выделенных по месту и способу образования, в инвентаре с относительной частотой соответствующих классов в текстах. И. Крамский исходит из предположения, что если бы употребление согласных разных классов было бы одинаковым в тексте, оно бы соответствовало доле этих согласных в фонемном инвентаре. Например, если в инвентаре имеется 8 различных взрывных согласных, 6 — различных фрикативных и 6 — назальных (т. е. 40, 30 и 30% всего инвентаря согласных), а в тексте употребляется 50% взрывных, 20% фрикативных и 30% назальных, то отклонения относительной частоты согласных в инвентаре от относительной частоты в тексте (здесь 10% для взрывных и 10% для фрикативных) дадут отклонение от равного употребления всех классов согласных. Положительная величина отклонения означает повышенное употребление, а отрицательная величина — недостаточное употребление.

По употреблению классов согласных соответственно способу образования исследованные 23 языка распадаются на три типа: 1) языки с повышенным употреблением взрывных и назальных; 2) языки с повышенным употреблением фрикативных и назальных (наименее многочисленный тип); 3) языки с повышенным употреблением назальных и плавных (наиболее распространенный тип).

По месту образования согласных языки делятся на четыре типа: 1) языки с повышенным употреблением лабиальных и переднеязычных; 2) языки с повышенным употреблением переднеязычных (наиболее распространенный тип); 3) языки с повышенным употреблением переднеязычных и палатальных; 4) языки с повышенным употреблением переднеязычных и велярных.

Следует отметить, что, насколько нам известно, работа Крамского представляет собой наиболее детальную попытку обследова-

[48] J. Kramský. A quantitative typology of languages. «Language and Speech», 1959, v. 2.

ния употребительности отдельных классов согласных в типологическом плане. Типология такого рода может дать много интересного для выяснения взаимного положения фонологических структур языков, однако само деление по классам, по-видимому, может быть усовершенствовано, с одной стороны, применением схемы бинарных различительных признаков, а с другой стороны — сопоставлением классификации согласных с классификацией гласных. Вероятно, целесообразнее применять не две шкалы классификации — по способу и месту, а одну, которой и являются универсальные фонологические признаки. Сам критерий повышенного и недостаточного употребления по сравнению с частотностью в инвентаре в принципе заслуживает одобрения; правда, если совмещать классификации гласных и согласных, то его применение может вызвать трудности, так как по сравнению с гласными частотность согласных в тексте меньше теоретической (в системе).

Другим важным (и методически более разработанным) направлением в квантитативной типологии является изучение употребительности фонем (и классов фонем) не в парадигматическом, а в синтагматическом плане. Выбирается единица анализа (слово или слог), устанавливаются ее основные структурные типы (виды последовательностей гласный + согласный... и т. п.), а затем подсчитывается, какого рода фонемы могут стоять на определенном месте в данном слово- или слоготипе. Таким образом составляются статистические модели основных структурных единиц языка.

В 1956 г. была опубликована работа И. Крамского о количественном анализе одно- и двусложных слов в английском языке [49], в которой указаны взаимоотношения звуков во всех позициях, а не только в соседних. Именно в этой работе И. Крамский впервые вводит понятия повышенного и недостаточного употребления фонем. По мнению автора, распределение звуков в словах подчиняется нескольким тенденциям: 1) предпочитающая тенденция, которая означает повышенное употребление некоторых фонем (или их классов) в определенной позиции в словотипе; 2) дискриминирующая тенденция, означающая недостаточное употребление некоторых фонем в определенной позиции в словотипе и 3) тенденция к равному распределению, означающая равное употребление всех классов фонем в одной или нескольких позициях в словотипе. Сравнивая поведение определенного класса фонем (например, взрывных) в двух или более различных позициях в словотипе, мы можем установить либо предпочитающую, либо дискриминирующую тенденцию. Первая означает усиление употребления фонем данного класса, а вторая — слабое их употребление по сравнению с другими позициями. Когда недостаточная употребительность определенных фонем в определенных позициях компенсируется их усиленным употреблением в другой позиции, можно говорить о компенсирующей тенденции.

[49] J. K r a m s k ý. On the quantitative phonemic analysis of English mono- and disyllables. «Časopis pro moderní filologie», 1956, v. 38, стр. 45—59.

Таким образом, идея пражской школы о глобальном изучении статистической структуры фонологии словаря получила у Крамского практическое выражение. Исследования в этом направлении были им продолжены в работе о статистике фонем в итальянском в одно-, дву- и трехсложных словах [50]. Здесь автор также опирается на данные словаря. Наличие предыдущих исследований подобного рода по английскому, персидскому (только односложные слова) [51], немецкому [52] и чешскому языкам [53] позволяет автору проделать, пусть ограниченное, типологическое строение статистической структуры слов в словаре этих языков. Привлекается также ограниченный материал по венгерскому и турецкому языкам.

Автор приходит к следующим выводам. Рассмотренные языки можно сгруппировать следующим образом согласно частоте отдельных классов согласных (по способу и месту образования): 1) итальянский и английский; 2) венгерский и турецкий; 3) персидский. Характерными чертами языков первой группы автор считает высокую частоту фрикативных, лабиальных и велярных для английского, высокую частоту палатальных для итальянского. Между обоими языками наблюдается согласие в частоте взрывных, назальных и альвеолярных. Подобным же образом обнаруживают близость в частоте согласных языки второй группы. Здесь, однако, для венгерского характерна большая употребительность палатальных, а для турецкого — большая употребительность фрикативных. Персидский противопоставлен всем остальным языкам по высокой частоте фрикативных, велярных и палатальных. Автор считает наиболее показательной гораздо меньшую частоту альвеолярных в персидском, чем в других языках.

Далее выделяются наиболее продуктивные словотипы. Среди односложных слов — это:

английский: CVC, CCVC, CVCC, CV, CCVCC
персидский: CVCC, CVC, CVVC, VCC, CV
турецкий: CVC, CVCC, VC
венгерский: CVC, CVCC, VC

Среди двусложных:

английский: CV́CVC, CV́CV, CV́CCVC, CV́CCV, CCV́CVC
итальянский: CV́CV, CV́CCV, CCV́CV, CV́CVV, CVV́CV

[50] J. Kramský. A quantitative analysis of Italian mono- di- and trisyllabic words. «Travaux linguistiques de Prague, I. L'Ecole de Prague d'aujourd'hui». Prague, 1964, стр. 129—144.

[51] J. Kramský. A Phonological analysis of Persian monosyllables. «Archiv orientální», 1947, v. 16, стр. 103—134.

[52] W. F. Twaddell. Combinations of consonants in stressed syllables in German. «Acta Linguistica», № 1. Kopenhagen, 1939, стр. 189—199; № 2, 1940, стр. 31—50.

[53] J. Vachek. Poznámky k fonologii českého lexika. «Listy filologické», v. 67, 1947, стр. 395—402.

Из этого сопоставления (особенно для двусложников) явствует тяготение английского языка преимущественно к закрытому типу слога, в то время как для итальянского характерно прямо противоположное — предпочтение открытых слогов.

Относительно этой работы можно было бы повторить все критические замечания, сделанные выше о методике статистики фонем в словаре. Сам автор сопровождает свое сравнение многочисленными оговорками о недостаточной репрезентативности избранного материала. По нашему мнению, в самой работе чувствуется стремление опираться на те факты, которые наиболее часты в речи: недаром для исследования взяты лишь одно-, дву- и трехсложники, интуитивно выделяемые как более частые, чем многосложные слова. С другой стороны, очевидно, что эти многосложные слова составляют в словаре не столь уж незначительный процент, чтобы ими можно было пренебречь совсем.

Таким образом, объективная логика исследования должна натолкнуть исследователей на то, чтобы и в изучении статистической структуры звуковых цепей опираться не на словарь (что кажется более простым), а на текст. Исследование текста и в этом случае должно уточнить выводы о более частой (или более редкой) встречаемости того или иного словотипа — ведь при анализе словаря весь ауслаут взят в сущности произвольно, правда, для такого нефлективного языка, как английский, это может быть и несущественным, но славянские языки должны быть представлены парадигмами слов, а не их лексикографическим репрезентантом. С другой стороны, исследование текста должно повысить в английском языке процент словотипов CV и VC, образующих столь многочисленные в тексте служебные слова.

Глубокое понимание квантитативной фонологической типологии мы находим в работах известных немецких фонологов П. Менцерата и В. Майер-Эпплера [54].

Согласно Менцерату, типология исследует механизм образования слогов и слов из последовательностей фонологических элементов в данном языке. Она пытается установить, что является типичным для данного языка в сфере звуков. Таким образом, при такого рода подходе типология исследует не разрозненные факты, относящиеся к изолированным уровням языка, а общий структурный принцип, организующий связь между всеми уровнями.

Для Менцерата типология начинается с вопроса, является ли словарь лишь механическим объединением, или он связан определенным структурным единством. В нашем случае, пишет Менцерат, словарь является системой, подлежащей исследованию, вскрытию. В том, что лексика является системой, не возникает сомнений.

---

[54] P. Menzerath, W. Meyer-Eppler. Sprachtypologische Untersuchungen, 1. «Studia Linguistica», Lund, 1950; P. Menzerath. Typology of languages. «Journal of the Acoustical Society of America», 1950, v. 22; P. Menzerath. Architektonik des deutschen Wortschatzes. Berlin, 1954.

Проблема состоит в том, чтобы найти критерии, определяющие специфическую картину структуры словаря. Автор показывает, как следует классифицировать словарные единицы, чтобы выявить эту структуру. Структура выявляется не на уровне смысла, а на конструктивном уровне: то, как из отдельных фонологических элементов конструируются слоги и слова, отражает многое в морфологии языка и т. п.

Каждое слово характеризуется дифференцированной формой, и, согласно критериям дифференциации, устанавливаются следующие типы: 1) словотипы (типы слов) — одно-, двусложные и т. п., например англ. speak, victim, kindliness; критерий: число слогов; 2) классы слов — одно-, дву-, трифонемные и т. п., например англ. are, two, three; критерий: количество фонем; 3) формотипы (типы форм), например англ. and, bit, three, три слова одного и того же типа (односложные) и класса (трифонемные), но разных форм: в каждом из трех случаев имеется другая аранжировка гласных и согласных; критерий: последовательность фонем.

Менцерат считает, что эти три критерия, вероятно, достаточны для классификации слов во всех языках с тем, чтобы обрабатывать их статистически.

Таким образом, используется лишь один чисто фонологический критерий — принадлежность фонемы к классу гласных или согласных, все другие фонологические признаки в этой классификации фонемных цепей не участвуют. Это обстоятельство уменьшает влияние того факта, что количество различных словотипов, формотипов и классов слов подсчитывается в словаре, а не в тексте: на уровне последовательности «гласный — согласный» статистическая структура фонологии словаря меньше отличается от статистической структуры фонологии текста, чем для других фонологических признаков.

Чтобы графически представить всю тотальность словаря каждого языка, Менцерат применяет весьма остроумный метод так называемых словарных пиков. Словарные пики представляют собой объемные трехмерные диаграммы, строящиеся по следующим координатам: горизонтальные оси: n — число фонем и z — число слогов, вертикальная ось — частота встречаемости различных словотипов, формотипов и классов слов, объединенных на этом графике. Таким образом, статистическая структура словаря в понимании Менцерата изображается как некоторое подобие горы или пика, вершина которого приходится на наиболее частый вид слова. Поскольку, по-видимому, наиболее частые формотипы могут не совпадать с наиболее частыми словотипами или классами слов, «ландшафт» может иметь несколько пиков и т. д. С помощью словарных пиков можно наглядно наблюдать многие характеристики слов в данном языке, например частотное распределение фонем в словаре, частоту различных типов слов согласно числу фонем и слогов и т. п. Идея словарного пика дает полезное графическое подспорье,

вскрывающее интересные структурные связи, пронизывающие разные уровни языка. К сожалению, подобный метод крайне трудоемок и требует особых навыков. Поэтому до сих пор работа Менцерата осталась в этом роде единственной.

Таким образом, типология Менцерата является типологией построения звуковых цепей на уровне «гласный — согласный».

Отношение между гласными и согласными в звуковой цепи — предмет исследования Э. Ньюмена [55].

Результаты анализа показывают, что, по мнению автора, способ упорядочения последовательности гласных и согласных, с одной стороны, удивительно постоянен для данного автора или вида материала. С другой стороны, наблюдаются значительные расхождения для различных языков или видов текста. Автор сравнивает одинаковые (небольшие по объему) отрывки из Библии на 11 языках. Он определяет два параметра: шаговую функцию автокорреляции (шаговая функция автокорреляции описывает степень зависимости между событием, входящим в последовательность в произвольное время t_0, и другим событием, которое входит в ту же последовательность и следует за первым в момент времени $t + t_0$), а также меру определенности (математический эквивалент негэнтропии).

Выясняется наиболее общая тенденция к взаимному чередованию гласных и согласных (т. е. к построению цепей типа CVCVCVCV...). Доказательством этого автор считает постоянные большие отрицательные значения коэффициентов автокорреляции для первого шага и постепенное уменьшение величины коэффициента для каждого последующего шага.

Согласно различным значениям коэффициента автокорреляции, Э. Ньюмен делит исследованные им языки на три группы: 1) языки с положительными значениями для второго шага (т. е. образующие цепи преимущественного типа CVC); 2) языки с отрицательным коэффициентом для этого шага (т. е. образующие цепи преимущественно типа VCC) и 3) языки с нулевым значением коэффициента, не отдающие предпочтения никакому типу.

К. Шеннон установил меру упорядоченности или структурирования. Она находится в обратном отношении к мере неопределенности, связанной с каждым символом. Согласно теории информации Шеннона, наибольшее количество информации может быть передано посредством системы с наименьшим количеством ограничений.

Э. Ньюмен определил меру определенности для каждого места в слове, начиная с нулевого (начало). Сравнение различных языков показало прежде всего, что имеются большие различия в определенных значениях, которые может принимать эта величина.

[55] E. B. Newmann. Pattern of vowels and consonants in various languages. «The American Journal of Psychology», 1951, v. 64, стр. 369—379. У нас подобные исследования проводили уже в 60-х годах Р. Г. Пиотровский и В. В. Шеворошкин.

Имеется также существенное различие в степени возрастания показателя от первого шага ко второму.

Интересна та часть статьи Ньюмена, где он сравнивает свои результаты с результатами, полученными Менцератом. Это сравнение можно суммировать следующим образом:

1) Менцерат брал слова на основе словаря, а результаты Э. Ньюмена основаны на связных текстах. Здесь методика Ньюмена предпочтительнее, так как учитывается влияние, оказываемое на фонологическую синтагматику грамматической структурой языка. Размеры текстов для целей Ньюмена не существенны, так как (мы это покажем в специальном разделе) на уровне последовательности гласных и согласных даже малые тексты являются случайной выборкой.

2) Словотипы Менцерата — односложники, в то время как материал Ньюмена охватывает все типы слов. По-видимому, это различие не столь существенно, так как (и это Ньюмен показал) зависимость символов друг от друга ограничивается пятью-шестью подряд, что приблизительно и составляет один слог.

3) Формотипы Менцерата описывают сочетания фонем, в то время как Ньюмен использовал материал в обычной орфографии. Поэтому различия оказались весьма велики: средняя длина односложных слов в немецком языке по Ньюмену 2,8 букв, по Менцерату — 3,8.

В этом пункте работа Ньюмена не выдерживает никакого сравнения с работой Менцерата. Приходится с сожалением констатировать, что весьма интересная методика, связанная к тому же с огромным количеством трудоемких вычислений, оказалась примененной к лингвистически не интересному материалу. В той мере, в какой алфавиты отражают фонологические системы языков, работу Ньюмена можно считать релевантной для фонологии. Однако среди исследованных языков такие, как английский и французский, чья орфография имеет довольно слабую связь с фонетикой, а также польский, в котором огромную роль играют разного рода диграфы и т. п.

Сам метод исследования упорядоченности звуковых цепей, предложенный Ньюменом, дает интересные результаты. С его помощью можно довольно четко вскрыть силлабическую структуру языка. Строго говоря, работа Ньюмена не является статистической. Мы решили изложить ее, во-первых, потому, что она относится к фонологической типологии и, во-вторых, потому, что это — одна из первых работ, в которой к анализу языка применена довольно изощренная математическая процедура. К сожалению, эта работа осталась пока без продолжения; вероятно (как и в некоторых других случаях) это объясняется большой трудоемкостью предварительных вычислений.

В 50-е годы интерес к квантитативной (фонологической) типологии возникает в американской лингвистике, особенно под влия-

нием работ Дж. Гринберга [56]. Представляется поучительной дискуссия, имевшая место в 1957 г. на страницах «Международного журнала по американистике». Дж. Пирс в своей работе предпринял обширную фонологическую типологию автохтонных языков Нового Света [57]. Эта типология основана на работе Ч. Фегелина о типологии языков американских индейцев [58], в которой фонологические системы классифицируются на основе числа смычных согласных рядов. Целью работы Пирса, по его словам, является статистическая проверка обоснованности такой группировки. В принципе работа Пирса лежит в русле идей относительно статистики фонем в системе (эти идеи стали к середине 50-х годов вполне общим местом, ср. высказывания Гринберга из его статьи 1957 г. о природе и применении лингвистической типологии: «Квантитативные атрибуты (лингвистических элементов.— *Д. С.*) можно рассматривать либо на основании встречаемости в системе (системные атрибуты), либо на основании встречаемостей в тексте»).

Однако Пирс придает этим идеям свое оригинальное направление. Он рассматривает не встречаемость в системе среди других объектов, но встречаемость в данной выборке языков. Иными словами, фонема получает частоту 1, если она встретилась в инвентаре лишь одного из 176 языков, и частоту 176 (наивысшую в данном случае), если она встретилась в инвентарях всех рассматриваемых языков.

Следует отметить, что со статистической точки зрения такой подход вполне закономерен, так как выборка (число рассматриваемых языков) достаточно велика. В этом случае понятие выборки имеет иной смысл, чем при статистике текста. Здесь встречаемость не является встречаемостью в речи, а представляет собой более «крупное» событие, поэтому число объектов в выборке может быть меньшим. Таким образом, идея Пирса в принципе не основывается на модели статистического порождения текста, как в работах по фоностатистике текста; у Пирса, следовательно, понятие «статистический» имеет иной смысл, чем в указанных работах.

В первой части своей статьи Пирс приводит частоты встречаемости всех согласных фонем в фонологических системах 176 языков. Выделяется всего 61 согласная фонема, причем самая частая *k* встречается в 173 языках, а самая редкая *k'* — всего в трех. Далее эти данные наносятся на прямоугольные координаты, где по абсциссе откладывается номер фонемы в списке по убывающей частоте,

---

[56] См., в частности: J. H. G r e e n b e r g. Essays in linguistics. Chicago, 1957; О н ж е. The nature and uses of linguistic typologies. «International Journal of American Linguistics», 1957, v. XXIII, № 2.

[57] Joe E. P i e r c e. A statistical study of consonants in New World languages.— IJAL, 1957, v. 23, стр. 36—45, 94—108.

[58] C. F. V o e g e l i n. Inductively arrived-at models for cross-genetic comparisons of American Indian languages. «University of California Publications in Linguistics», 1954, 10, стр. 27—45.

а по ординате — ее частота. Получается некоторое эмпирическое распределение. Делается предположение о том, что если бы это распределение было полностью случайным, оно было бы, по крайней мере, непрерывным. В действительности эмпирическое распределение обнаруживает два разрыва. Из этого делается вывод, что это распределение образуется из смешения трех различных распределений, отражающих не только различия в частоте встречаемости фонем в инвентарях исследуемых языков, но и значимые различия между соответствующими группами фонем на собственно фонологическом уровне.

Первое распределение автор аппроксимирует как линейное. (Линейность означает, что различие в относительной частоте встречаемости двух соседних по рангу согласных постоянно. Экспериментальные относительные частоты: 0,99, 0,98, 0,97, 0,94, 0,93, 0,92, 0,89, 0,88, 0,86; они относятся к фонемам *ktnsmpyhw*. Естественно, что нет строгой линейности, но приблизительно это можно считать верным). Таким образом, первое распределение включает в себя согласные, которые можно характеризовать как идентифицируемые наименьшим количеством дифференциальных признаков. Пирс называет эту систему типичной консонантной системой для языков Нового Света. Он подчеркивает, что ее выделение основывается исключительно на формальных основаниях: 1) разрыв в распределении между *w* (последняя фонема первой системы) и *tš* (первая фонема второй системы) — 15%, т. е. больше, чем между первой и последней фонемами первой системы; 2) верхнее распределение приблизительно линейно, в то время как второе, видимо, аппроксимируется параболой или гиперболой.

Подобным же образом вторая система, куда входят *tš*, *l*, *ts*, *b*, *x*, *ʔ*, *d*, *g*, *r*, отделяется от третьей: разрыв между ними составляет 20%, в то время как максимальное падение относительной частоты между фонемами II группы составляет 7%; таким образом, наблюдается вполне отчетливое стремление согласных группироваться в фонологически значимые группы, разрывы в частоте между которыми заведомо превышают аналогичные разрывы внутри групп. Вид распределения в третьей группе также отличается от первых двух. Это распределение можно аппроксимировать как логарифмическое.

Выводы Пирса, изложенные выше и содержащиеся в первой части его статьи, заслуживают пристального внимания. Представляется не случайным, что между согласными I и II группы имеется столь четкое отношение по составу фонологических признаков: согласные II группы можно представить как согласные I группы (условно «простые») плюс один различительный признак: *bdg* = = *ptk* + «звонкость», *tš* = *t* + «компактность», *ts* = *t* + «яркость», *x* = *k* + «непрерывность», *ʔ* = *y* + «абруптивность» и т. п. Столь четкое расслоение системы согласных по встречаемости фонемы в инвентаре соответствует, как мы помним, правилу Ципфа,

согласно которому чем «сложнее» звук, тем реже он встречается в тексте. Совпадение классификаций, основанных на столь различных основаниях, кажется тем более значимым.

Дополнительное соображение о полезности подобного обследования фонологических систем для типологии заключается в том, что системы типа системы I у Пирса (согласные, встречающиеся в 90% обследованных языков), по-видимому, отражают некоторую лингвистическую реальность.

За последнее время в ряде работ у нас[59] появилось понятие теоретико-множественного произведения фонологических систем языков, которое отчасти совпадает с системой I по Пирсу. Правда, в указанных работах ТМП выделяется только в группе родственных языков, которые насчитывают значительно меньше языков, чем 176, исследованных Пирсом, поэтому в них в ТМП входит больше членов, чем в систему I Пирса. Сравним три минимальные системы консонантизма: систему Пирса для языков Нового Света (I), систему, выделяемую группой авторов для славянских языков (II), и систему ТМП для дардских языков (III):

I	II	III
p t k	*p b t d kg*	*p b t d td kg*
s	*s z*	*c č*
m n	*m n*	*s z š*
w y h	*n'*	*m n*
	r l j	

Сравнение подобных «минимальных» систем наглядно суммирует наиболее существенные признаки каждого языкового ареала: наличие корреляции «палатализованность—непалатализованность» даже в минимальной системе для славянских языков, наличие ряда церебральных в дардском ареале и существенное различие в строении системы сонантов в минимальных системах: в минимальной системе для языков Нового Света назальным противопоставляется ряд «глайдов» (*w y h*), в славянских — ряд плавных, а в дардских в ТМП входят лишь одни назальные.

Вместе с тем подобное сравнение показывает недостатки типологии Пирса: она основывается лишь на сравнении количества рядов смычных, а не на изучении всей консонантной системы — существенные типологические корреляции наблюдаются и за пределами системы смычных. Если первая часть статьи Пирса была

---

[59] См., в частности: М. И. Лекомцева, Д. М. Сегал, Т. М. Судник, С. М. Шур. Опыт построения фонологической типологии близкородственных языков. «Славянское языкознание». М., 1963; В. Н. Топоров. Предварительные материалы к описанию фонологических систем дардских языков. «Лингвистические исследования по общей и славянской типологии». М., 1966.

посвящена методике выделения «общей части» фонологических систем сравниваемых языков, то вторая часть посвящена собственно типологии. Здесь Пирс указывает, что из 176 языков в 66 — один ряд смычных (тип I), для этого типа максимальное количество согласных фонем — 28, минимальное — 9 и среднее количество согласных — 14,5 на язык. Такое же количество языков (66) имеет два ряда смычных (тип II) — максимальное число согласных фонем — 33, минимальное — 13, среднее — 20,7 на язык. В 38 языках три ряда смычных (тип III) — максимальное количество согласных фонем — 46, минимальное—18, среднее число согласных в языке этого типа — 29,3. В 5 языках — 4 ряда смычных (тип IV), при этом максимальное количество согласных фонем — 36, минимальное — 21, среднее — 28,2. Далее автор показывает, что каждый тип представлен группой языков, которые дают параллельное или сходное распределение, причем распределение для каждого типа отличается от соответствующего распределения для других типов. Для того, чтобы подчеркнуть гипотезу, что подобная группировка объединяет в каждом типе консонантные системы с общими фонологическими чертами, автор анализирует системы каждой группы, устанавливая согласные, репрезентативные для языков каждого типа.

Подход Пирса к типологии и результаты его исследования подверг критике Сол Сапорта [60]. Наиболее существенное в критике Сапорты — это его анализ исходных данных, которыми пользовался Пирс. Зачастую это были фонетические, а не фонологические описания, поэтому иногда решения о фонологической значимости или незначимости того или иного признака исходили из совершенно произвольных оснований (например, в языке майду имплозивные *p̀ d̀ g̀* были зачислены в разряд просто звонких, в то время как неясно, какой именно признак — «звонкость» или «имплозивность» является действительно релевантным). Поэтому Сапорта делает замечание о том, что прежде чем начать сравнивать языки по методу Пирса надо сравнить фонологические системы сами по себе, а не только используемые символы. Русское и английское (*p*) не могут быть приравнены друг к другу, так как при идентификации русского (*p*) используется признак «палатализованность—непалатализованность», который не существен для английского (*p*). Правда, это возражение можно игнорировать в тех случаях, когда сопоставляемые фонемы различаются по признаковому составу, но отсутствует интерферирующий член. Например, в немецком языке (*t*) и (*d*) противопоставляются как сильный и слабый, а во французском как глухой и звонкий, однако поскольку в обоих языках нет «других» *t* и *d*, то указанные фонемы можно считать равнозначными в соответствующих системах.

---

[60] Sol Saporta. Methodological considerations regarding a statistical approach to typologies.— IJAL, 1957, v. 23, стр. 109—113.

Во всяком случае, Сапорта прав в том, что необходимо точнее устанавливать отношения фонем, чем это сделал Пирс. Второе замечание Сапорты сводится к тому, что Пирсу в первой части статьи не удалось показать, что его эмпирическое распределение отличается от случайного. Сапорта утверждает, что если бы использовались данные более точного фонологического анализа, то распределение получилось бы непрерывным (в частности, уменьшилось бы количество языков, содержащих *w* и *y*, так как они, по мнению Сапорты, перешли бы в *u* и *i*). Нам представляется, что это замечание должно быть отклонено, поскольку разрывы в частоте между группами настолько велики и эмпирические распределения настолько отличаются друг от друга, что не требуется сложных статистических тестов, чтобы это установить.

Более существенным представляется возражение против использования критерия количества рядов смычных как первичного при типологическом анализе. Ясно, что когда консонантные системы классифицируются по количеству смычных, число согласных на язык будет больше в языках с двумя рядами смычных, чем в языках с одним рядом. Таким образом, оба признака являются зависимыми друг от друга и один из них явно лишний. Далее Сапорта считает, что если в языке используется голос для установления второго ряда смычных, тот этот новый различительный признак явно не будет ограничиваться лишь смычными. Язык, в котором два ряда смычных, вероятно, будет иметь и два ряда фрикативных. Сапорта также упоминает о том, что смычные могут дифференцироваться не только по способу, но и по месту образования.

Еще до дискуссии Сапорты и Пирса Р. Уэллс отмечал [61], что классификация является естественной, когда каждый класс характеризуется признаками, априори не подразумеваемыми определяющим его качеством. Поэтому следует ожидать корреляции между числом согласных в системе и количеством смычных, а не между количеством рядов взрывных и количеством мест артикуляции у смычных. Классификация Фегелина, принятая Пирсом, основана исключительно на инвентаре. Однако системы могут различаться и типами возможных консонантных сочетаний, нагрузкой различительных признаков и т. п. Очень существенным является и то, что Пирс совершенно не учитывает фактов вокализма, в то время как можно ожидать значимых соответствий между типами вокалической и консонантной систем. Вообще проблематично, можно ли строить типологию лишь на том факте, что в языке один или больше рядов смычных в инвентаре согласных. Как уже отмечалось, этот факт зависит от общего числа согласных и именно вследствие этой зависимости не может служить надежным типологическим критерием.

---

[61] R. W e l l s. Archiving and language typology.— IJAL, v. 20, 1954.

* * *

В 60-е годы развитие лингвистической статистики приняло настолько бурный характер, что невозможно учесть все конкретные исследования по самым различным языкам мира, которые появились в печати.

Шестидесятые годы характеризуются прежде всего тем, что наряду с конкретными работами, в которых методы лингвистической статистики применяются для решения определенных задач, появляются обобщающие теоретические труды, в которых делается попытка дать теоретическое обоснование применения статистических методов с целью модифицировать их для специально лингвистических целей.

Мы имеем в виду труды Пьера Гиро [62] и Густава Хердана [63]. Из предыдущего обзора можно заключить, что до появления работ Гиро и Хердана все работы по статистической фонологии распадаются на две группы: работы, интерпретирующие экспериментальные количественные данные с точки зрения фонологии, и работы, применяющие отдельные методы и приемы математико-статистического аппарата к лингвистическому материалу. Значение книг Гиро и Хердана в том, что в них предпринято теоретическое осмысление самого метода и аппарата статистики с точки зрения лингвистики (а не наоборот, как в более ранних работах,— нахождение в лингвистической практике и теории релевантных для статистики моментов). Кроме того, эти труды представляют собой первое систематическое изложение статистики специально для лингвистов, при этом лингвистика понимается как целое, а не как лишь одна часть ее (например, только изучение словаря или фонологии).

Рассмотрим книгу П. Гиро «Проблемы и методы лингвистической статистики». П. Гиро — не математик, а филолог, автор интересных работ по теории стихосложения и стилистики. Поэтому его работа ориентируется на читателя, более заинтересованного результатами статистического анализа, чем его теоретическими обоснованиями. По нашему мнению, к наиболее удачным частям книги (равно как и всей работы Гиро в области лингвистической статистики) относятся главы, посвященные конкретному статистическому анализу структуры лексики («Уравнение Эсту—Ципфа и статистические характеристики словаря» — глава, повторяющая

---

[62] См. прежде всего: P. G u i r a u d. Problèmes et méthodes de la statistique linguistique. Paris, 1960. Гиро также составил превосходную библиографию по лингвистической статистике, включающую почти 2500 названий: «Bibliographie critique de la statistique linguistique». Utrecht, 1954.

[63] Gustav H e r d a n. Language as Choice and Chance. Groningen, 1956; О н ж е. Type-token Mathematics. The Hague, 1960; О н ж е. The Calculus of Linguistic Observations. The Hague, 1962; О н ж е. Quantitative Linguistics. London, 1964; О н ж е. The Advanced Theory of Language as Choice and Chance. Berlin, 1966.

основные положения книги Гиро на ту же тему [64], а также главы «Эволюция стиля Рембо» и «Фонетическая структура стиха»). Мы не будем здесь подробно анализировать эти главы, так как их ценность заключается именно в конкретности, филологической значимости, и понадобилось бы просто перевести их целиком, чтобы дать представление о подходе автора. Кроме того, эти разделы книги касаются в основном лексического уровня, который мы не включаем в настоящий обзор. Укажем лишь, что в отличие от многих работ по применению статистических методов к анализу словаря и стиля работа Гиро отдает статистике роль инструмента, а не цели. Гиро всюду ставит осмысленные и весьма интересные вопросы о динамике развития стиля автора, о датировке отдельных вещей, о стилистической иерархизации словаря писателя (или текста) в зависимости от частоты слов и т. п.

Иными словами, работа Гиро дает много не только для лингвистов, занимающихся статистикой, но прежде всего для литературоведов, филологов и специалистов по стилистике, интересующихся исключительно материалом, к которому применяются статистические методы.

В книге Гиро имеется специальная глава, в которой рассматривается вопрос о статистических методах в фонологии. Здесь Гиро затрагивает один аспект проблемы, а именно—интерпретирует изложенную нами выше теорию фонетической частотности Дж. Ципфа в свете введенного им понятия оценки отклонения (écart-reduit) s/σ, где s — абсолютное отклонение от среднего значения, а σ — среднее квадратичное отклонение. По сути дела эта глава является лишь слегка модернизированным изложением теории Ципфа, никаких новых лингвистических данных здесь не содержится. Гиро использует данные о частотности фонем, содержащиеся в работах Ципфа (эти данные, как мы отмечали, во многом неточны), и совершенно не затрагивает новых работ и материалов, появившихся после Ципфа. Тем не менее эта глава интересна тем, что положения Ципфа связываются здесь с некоторыми идеями теории информации. Согласно Б. Мандельброту [65] уравнение Ципфа $p_r = к \cdot r^{-\gamma}$ постулирует существование экономии при передаче сообщений в виде слов, экономии, заключающейся в том, что информационное содержание слов, определяемое их вероятностью, пропорционально их сложности.

Из этого делается заключение о том, что если в языке фонемы встречаются с большой стабильностью, то их вероятность отражает экономию, присущую языковому коду и пропорциональную их сложности, иными словами, чем чаще встречается фонема, тем она «проще», и чем реже — тем сложнее.

---

[64] P. Guiraud. Charactères statistiques du vocabulaire. Paris, 1954.

[65] См., в частности: M. Mandelbrot. Structure formelle des textes et communication. «Word», 1954, № 10, стр. 1—27.

До сих пор рассуждение Гиро совершенно совпадает по направлению с рассуждением Ципфа в его книге «Относительная частота как определяющий фактор фонетического изменения».

Далее Гиро касается вопроса о сложности фонемы. Здесь он уже ближе к современному пониманию, чам Ципф. Для Гиро сложность не является столь очевидным понятием, как для Ципфа. Сложность «предстает как результат воздействия столь большого комплекса причин, что при современном состоянии исследований нет никакой возможности расчленить этот комплекс на составные элементы» [66]. И далее: «Следует признать, что именно это сложное качество, не поддающееся анализу и воспринимаемое глобально, регулирует частотность фонемы» [67].

Гиро выделяет три функции фонологической системы. Диакритическая функция, с помощью которой осуществляется распознавание знака среди других знаков, обеспечивается формой фонологической системы в том смысле, в каком этот термин употребляется в структурной лингвистике. Система противопоставлений фонем получает название фоно-диакритической системы. Далее выступает информационная функция, обеспечивающая идентификацию знака в условиях максимальной экономии. Эта функция обеспечивается существованием системы стабильных частот. Соответствующая система называется фоностатистической. Третья функция—семантическая, связывающая форму знака со значением, по мнению Гиро, не влияет на организацию фонологической системы и поэтому не рассматривается.

Автор вслед за Ципфом выдвигает понятие равновесия системы: «Равновесие фонологических систем — в том смысле, в каком этот термин употребляется в структурной лингвистике,— т. е. фонодиакритических систем — опирается на четкую структуру, которую образует совокупность фонем, являющихся пучками различительных признаков, по которым фонемы противопоставляются в оппозиции... Это равновесие соответствует определенной функции. С помощью этого равновесия обеспечивается лучшее диакритическое функционирование системы; равновесие позволяет получить с помощью минимального количества фонологических признаков максимальное количество оппозиций... Понятия равновесия и диакритической функции оказываются удобным инструментом при анализе фонологической эволюции. В частности, обнаруживается, что стабильность фонемы или корреляции пропорциональна их диакритической упорядоченности и их интеграции в системе; с другой стороны, система стремится восстановить свое равновесие в тех случаях, когда оно нарушено» [68].

Мы помним, что совершенно аналогичные рассуждения были у Ципфа, который выдвинул гипотезу о том, что состояние равнове-

[66] P. Guiraud. Charactères statistiques du vocabulaire, стр. 97.
[67] Там же, стр. 98.
[68] Там же, стр. 99.

сия (или нарушения равновесия) системы отражается в частоте фонемы. Гиро целиком заимствует это предположение, и в его изложении эта мысль выступает как постулат о зависимости фонодиакритической системы от системы фоностатистической.

Гиро весьма высоко оценивает вклад Ципфа в развитие теоретической фонологии и лингвистической статистики. Он пишет: «Ципф выдвинул гипотезу о том, что частота пропорциональна усилию, необходимому для производства фонемы; чаще всего употребляются наиболее простые фонемы; функцией системы является экономия усилий; Ципф строит всю свою теорию вокруг «принципа наименьшего усилия», куда он включает также проблему частоты слов.

К сожалению, Ципф умер слишком рано и не успел окончательно отшлифовать свою гипотезу; разумеется, в том виде, в каком она изложена, она изобилует наивными предположениями и натянутыми интерпретациями, которые еще предстоит проверить. Поэтому эти работы не без основания подвергались критике. Однако не следует забывать, что Ципф был пионером, преждевременно ушедшим из жизни в тот момент, когда прогресс физики и теории коммуникации, возможно, дал бы в его распоряжение соответствующие критерии для уточнения и разъяснения его гипотез.

В принципе можно утверждать, что «теория наименьшего усилия» Ципфа остается ценным постулатом, который надлежит уточнить и определить его границы, стремясь при этом очистить его от неясности и тумана, в который эта теория облечена вследствие особенностей изложения автора.

Понятие информации позволяет нам сегодня повторить анализ Ципфа с большей строгостью и точностью, опираясь на более детальный статистический анализ» [69].

Нам остается посмотреть, как Гиро проводит фоностатистический анализ и насколько его метод более точен.

В дополнение к таблицам Ципфа, показывающим частоту фонем в 17 языках (в число которых включены не уточненные nord-russe, sud-russe и wende, по-видимому, русский, украинский и, видимо, один из лужицких языков (?), Гиро приводит средние величины частоты по каждой из фонем для 17 языков, среднее квадратичное отклонение и коэффициент вариации, а также, отдельно, оценки отклонения (écarts-reduits) для взрывных фонем, для подклассов внутри взрывных и фрикативных и для *m*. Анализ средних значений не вносит ничего нового в выводы Ципфа и не подкрепляет их никаким статистическим аппаратом, поскольку средние значения зависят от подбора языков (а он, в общем, случаен) и от вида выборки в каждом языке. Вообще представляется, что выводить среднее значение частоты фонем для нескольких языков, особенно если эти языки не связаны никакой общностью

[69] P. Guiraud. Указ. соч., стр. 99—100.

(территориально, генетически или типологически),—операция достаточно рискованная и мало осмысленная. Она возможна лишь как способ описания расхождения между языками в определенном признаковом пространстве, но не как определение реальной величины, к которой стремится частота фонемы в каждом языке (а именно так трактует эту величину Гиро, незаметно для себя спускаясь с уровня моделирования на уровень установления причинно-следственных зависимостей). Поэтому выводы о том, что «глухие встречаются чаще звонких в отношении 7 к 3», а также о том, что «апикальные — наиболее частые из всех подклассов взрывных», изложенные в столь общем виде для столь малого количества языков (при этом лишь для части системы — взрывных шумных), представляют мало интереса, и к ним можно отнести сказанное выше по поводу работы Бурдона, опубликованной в прошлом веке.

Примерно так же обстоит дело и с анализом отклонений, поскольку величины среднего квадратичного отклонения и коэффициента вариации столь же зависят от подбора языков и выборки, как и средние значения частот. Впрочем, данные величины выражают эту зависимость не столь непосредственным образом, как среднее значение, поэтому их можно хотя бы в первом приближении использовать для сравнения фонологических систем языков. Гиро считает, что коэффициент вариации (v) позволяет оценить стабильность фонемы или класса фонем: чем меньше его значение (т. е. чем меньше варьируется частота данной фонемы или класса в пределах данной выборки языков), тем более стабильной можно считать частоту этой фонемы. Получается, что взрывные все вместе более стабильны, чем фрикативные ($v = 0{,}12$ против $v = 0{,}19$). Это, по мнению Гиро, отвечает интуитивному предположению о том, что взрывные фонемы являются фонемами стабильными в отличие от фрикативных. Звонкие гораздо более нестабильны ($v = 0{,}24$) по сравнению с глухими ($v = 0{,}12$). Внутри класса взрывных наиболее стабильным является подкласс дентальных ($v = 0{,}18$). Самую меньшую стабильность показал класс заднеязычных фрикативных ($v = 0{,}47$).

В принципе изучение стабильности частоты фонем является крайне важным, и заслугой Гиро является то, что он первым предпринял практические шаги в этой области; однако предложенной методике не хватает по крайней мере двух моментов, чтобы стать настоящей строгой методикой. Во-первых, как уже отмечалось, подбор языков должен быть более обоснованным, желательно привлечение большого числа разнообразных по фонологической структуре языков, и главное, частотные данные должны быть достоверными. Во-вторых, следует иметь какой-то способ установления статистической значимости расхождений между параметрами: является ли расхождение между $v = 0{,}12$ и $v = 0{,}19$ значимым или оно могло возникнуть в результате случая? Все это очень важ-

ные вопросы и они вполне могли быть хотя бы поставлены в работе Гиро. То, что этого нет, снижает ее ценность.

Ниже мы предложим собственный способ оценки стабильности частоты фонем внутри *одного* языка (что кажется более обоснованным), и тогда некоторые выводы Гиро можно будет уточнить. Отметим, что общее представление о взрывных как о более стабильных фонемах можно извлечь не только из статистических данных, но и из некоторых типологических представлений (треугольник *p t k* имеется почти во всех языках мира, ср.: Н. С. Трубецкой. Основы фонологии, стр. 142), а также из данных онтогенеза. Статистические данные могут эти соображения подкрепить.

Таким образом, в области сравнения средних значений и отклонений методика Гиро, к сожалению, не дала всего того, что можно было бы ожидать от нее, если бы материал, к которому она применялась, был более основательным.

Далее Гиро затрагивает интересный вопрос о соотношении реальных частот с частотами теоретическими. Теоретические частоты получаются им сразу для всех 17 языков (по средним значениям), исходя из предположения, что реальная фонема образуется как наложение ряда фонологических признаков одного на другой. Теоретическая частота получается как произведение частот этих признаков. Иными словами, Гиро исходит из предположения о том, что отдельные признаки выступают как независимые друг от друга события; кроме того, имеется другое существенное допущение, что признаки существуют как единицы, отдельные от фонем. В принципе против обоих допущений трудно что-либо возразить на априорных основаниях, особенно если трактовать признаки не как реальные физические сущности, а как абстрактные символы. При этом, однако, необходимо иметь некоторый способ, который бы позволял определять частоту признаков независимо от частоты фонем. В рассматриваемой работе дело обстоит не так. Частота признаков выводится из частоты фонем (частота признака «глухость», например, определяется как сумма частот всех отдельных глухих фонем), и это совершенно естественно, ибо иного способа получить эту частоту нет. Однако поскольку исходные цифры взяты из частот фонем, то совершенно очевидно, что при перемножении соответствующим образом подобранных цифр мы получим величину, очень близкую к исходной. К тому же конкретный способ выбора частот тоже вызывает вопросы. Каждая фонема у Гиро определяется лишь двумя признаками: «звонкость» — «глухость» и место образования (дентальность, лабиальность, велярность). Частота звонкости или глухости берется в отношении ко всем фонемам языка (соответственно 5,43 и 12,77%), в то время как частота места образования берется относительно только взрывных фонем (соответственно 52, 22 и 26%). Можно было взять наоборот — частоту звонких и глухих в отношении к взрывным, а частоту мест образования

в отношении ко всем фонемам. Мы проверяли это, и цифры оказались неизменными.

Однако в реальных языках фонемы зачастую определяются более чем двумя признаками («звонкость» и «место»), иногда на них может накладываться признак тембра (церебральные или палатальные и т. п.). В этом случае крайне трудно выбрать такую аранжировку частот, чтобы при перемножении получить величину того же порядка, что и частота фонемы. Поэтому метод получения теоретических частот, предложенный Гиро, сам по себе не дает достоверных результатов для реальных языков, а лишь для специально подобранных средних значений, относящихся не к целой фонологической системе, а к ее упрощенному фрагменту.

В заключение своей работы Гиро излагает вводимый им способ оценки состояния фонологической системы. Этот метод основывается на том, что в таблице сравниваемых языков для каждой фонемы данного языка вычисляется оценка отклонения (écart-reduit). Автор предлагает считать, что если величина оценки колеблется около 1, то данная фонема находится в равновесии; если эта величина заключена между 1 и 1,5, то фонема начинает терять равновесие; если оценка равна 1,5—2, то фонема потеряла равновесие; если оценка превысила 2, то имеется патологически ненормальное состояние. Гиро указывает, что эти границы оценки являются статистической интерпретацией введенного Ципфом понятия допустимого порога, которое до сих пор было подкреплено лишь интуитивными соображениями. По мнению Гиро, подобные оценки должны сыграть большую роль в диахронической фонологии, где отсутствие количественных параметров лишает рассуждения точности и доказательности (в качестве примера Гиро ссылается на некоторые места из книги А. Мартине «Экономия фонетических изменений», которую в целом он оценивает высоко и считает, что положения Мартине подкрепляют теорию Ципфа).

Применение параметра écart-reduit для изменения равновесия системы иллюстрируется Гиро на тех же примерах, которые были использованы Ципфом,— элиминация конечного *m* в латинском (écart-reduit = + 2,80) и озвончение интервокальных глухих в некоторых романских языках.

Достигаются ли какие-нибудь результаты дополнительно к тому, что уже ранее было показано Ципфом? Нам кажется, что в данном конкретном случае сами цифры частоты были настолько велики, что не требовалось дополнительных критериев для доказательства того, что *m* в латинском языке нарушает равновесие системы. Более того, оценка отклонения сама по себе зависит от количества сравниваемых языков и от их фонологических систем, в то время как цифра частоты от этого не зависит. По-видимому, следовало бы избрать другие примеры для иллюстрации действия этого параметра, поскольку устранение латинского *m* уже получило свое объяснение в работе Ципфа.

Если давать общую оценку фонологической части книги Гиро, то следует отметить, что реальные результаты работы оказались ниже уровня обещаний, которые сам автор дал в начале главы. Материал Ципфа достаточно интерпретирован самим Ципфом, и привлечение немного более изощренной статистической техники не очень меняет дело. При всем этом статистические параметры, введенные Гиро, достаточно просты и удобны в обращении. Поэтому необходимо было бы применить их к более сложному материалу. Только тогда можно будет сделать окончательный вывод о нетривиальности (или тривиальности) его методики.

В работе Гиро подытоживаются идеи и проблемы всего предшествующего развития лингвистической статистики. Суммируя сказанное о конкретных работах в этой области начиная с Ципфа, можно выделить следующие основные идеи и наблюдения фонологической статистики:

а) постулируется стабильность частоты фонем для данного языка:

б) «более простые» фонемы встречаются чаще, чем «более сложные»;

в) постоянство частот делает возможной квантитативную типологию как на уровне слова, так и на уровне фонологической системы.

Квантитативная фонология развивается все время в круге этих идей, привлекая для анализа новые языки и совершенствуя применяющиеся статистические методы. Однако это развитие представляется нам несколько односторонним по следующим причинам:

а) крайне ограничен и недостаточен круг материалов, используемых в фонологической статистике; многие серьезные работы основываются на одних и тех же устарелых данных Ципфа. Нет исчерпывающих монографических описаний статистики фонологического уровня для подавляющего большинства языков (исключения крайне немногочисленны, см. начало этой главы о статистике новоиндийских языков);

б) бедность материалов сочетается с крайней фрагментарностью статистических приемов, используемых при анализе. Нет единой теории лингвистической статистики, не показано, чем применение статистических методов в лингвистике отличается от применения этих методов в других областях. Совершенствование методов идет лишь по линии обращения к новым разделам математической статистики, без построения единого аппарата лингвистической статистики.

Упомянутые выше работы Г. Хердана коренным образом изменили состояние лингвистической статистики. Если Дж. Ципф явился первым, кто обратил внимание на теоретико-лингвистическое значение частотности лингвистических элементов, то Г. Хердан по сути дела создал заново весь концептуальный и методический аппарат современной лингвистической статистики. Основные теоретико-статистические положения настоящей работы возникли как результат усвоения некоторых идей Хердана и стремления проверить их релевантность.

Глава вторая

ПРОБЛЕМА СТАБИЛЬНОСТИ ЛИНГВИСТИЧЕСКИХ ЧАСТОТ И СОВРЕМЕННОЕ СОСТОЯНИЕ ЛИНГВИСТИЧЕСКОЙ СТАТИСТИКИ

§ 1. Некоторые необходимые понятия

Для того, чтобы иметь возможность обсуждать интересующие нас вопросы статистики, введем некоторые необходимые математические понятия. Прежде всего мы будем пользоваться такими понятиями, как событие, случайная величина, теоретическая и эмпирическая вероятность, теоретическая и эмпирическая функция распределения, генеральная совокупность и выборка. Все эти понятия выработаны в теории вероятностей и математической статистике, и применение их к лингвистическим объектам связано с целым рядом допущений (о массовости, повторяемости и случайности), о которых говорилось во Введении.

Как правило, статистическое распределение характеризуется типом функции распределения и параметрами: средним значением (математическим ожиданием) и дисперсией случайной величины.

Среднее значение, или математическое ожидание $M\xi$ случайной величины ξ, принимающей лишь конечное число значений $t_1,\ldots t_n$, определяется как сумма произведений этих значений и соответствующих им вероятностей:

$$M\xi = \sum t_k p_k .$$

Из определения [1] ясно, что среднее значение заключено между наименьшим и наибольшим возможными значениями случайной величины ξ. Среднее значение можно также определить как центр тяжести возможных значений t_k с весами p_k.

Очень важно оценить величину разброса значений случайной величины ξ от среднего значения $M\xi$. Таким параметром является

[1] См.: Б. Л. ван дер Варден. Математическая статистика. М., 1960.

дисперсия $D\xi$, которая вычисляется как сумма квадратов отклонений индивидуальных значений случайной величины от среднего значения:

$$D\xi = \sum (\xi - M\xi)^2.$$

Естественно, что индивидуальные значения случайной величины могут быть больше или меньше M, поэтому индивидуальные отклонения могут иметь положительный или отрицательный знак. Чтобы избавиться от отрицательного знака, величину отклонения возводят в квадрат: затем суммируют по всем значениям случайной величины. Обычно в качестве параметра распределения используются $\sqrt{D\xi}$ или среднее квадратичное отклонение σ.

Функция распределения — это способ упорядоченного представления соотношения значений случайной величины и соответствующих этим значениям вероятностей. Она строится таким образом, что на абсциссе откладывается вариационный ряд значений случайной величины, а по ординате откладываются накопленные значения вароятности, так, что первому (наименьшему) значению абсциссы соответствует его вероятность p_1, второму значению абсциссы — сумма вероятностей $p_1 + p_2$, третьему значению абсциссы — $p_1 + p_2 + p_3$, а последнему (наибольшему) значению Σp_i, т. е. 1.

В случае эмпирической функции распределения теоретическое среднее значение и теоретическая дисперсия оцениваются по выборочному среднему значению и выборочной дисперсии.

Нормальное распределение характеризуется тем, что наибольшее значение интервала вероятности приходится на интервал вокруг среднего значения случайной величины: кривая плотности вероятности симметрична и имеет «колоколообразную» форму и, наконец, это распределение слагается из суммы воздействий большого количества факторов, влияние каждого из которых само по себе ничтожно.

Кривая плотности распределения Пуассона — асимметрична и здесь наибольшее значение интервала вероятности приходится не на интервал вокруг среднего значения случайной величины.

Оба распределения являются теоретическими, эмпирические распределения, с которыми имеет дело статистика в своих приложениях, как правило, могут быть сведены к нормальному распределению, или распределению Пуассона (или некоторым другим теоретическим распределениям). Существуют также различные модификации этих распределений, введенные с целью добиться лучшей аппроксимации экспериментальных данных. Большинство статистических процедур либо выполняется в предположении нормального распределения, либо позволяет проверить степень приближения экспериментальных данных к теоретическим данным, диктуемым тем или иным законом.

Процедуры проверки принадлежности данного эмпирического распределения к нормальному распределению, или распределению Пуассона, выполняются при помощи разного рода статистических критериев значимости. Прикладная статистика имеет дело с самыми различными статистическими критериями; одним из наиболее распространенных является критерий χ^2. Критерий χ^2 применяется тогда, когда по наблюденным частотам нужно проверить некоторую гипотезу относительно вероятностей. Общая формула этого критерия такова:

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - np_i)^2}{np_i},$$

где x_i — наблюденная частота; p_i — теоретическая вероятность (которая предполагается априори известной!); n — число наблюдений (иначе — объем выборки).

Эта формула, как мы видим, включает в себя заранее известную теоретическую вероятность p_i. Во многих случаях эта вероятность не известна заранее, а ее следует определить из предположения о том, что экспериментальное распределение следует некоторому теоретическому закону.

Таким образом, если мы делаем предположение о том, что наше исходное распределение подчиняется нормальному распределению, или распределению Пуассона, то мы по специальным формулам вычисляем значения p, затем подставляем их в формулу критерия χ^2, вычисляем значение χ^2 для данного x и проверяем получившийся результат по специальным таблицам. Если величина χ^2 оказывается больше определенного предела, то гипотезу относительно вероятностей следует отвергнуть.

Критерий χ^2 может применяться не только для проверки принадлежности экспериментального распределения определенному теоретическому распределению, но и для других целей.

Разумеется, мы крайне приближенно описали здесь работу этого критерия. Ниже мы подробнее коснемся его устройства и применения. В данном случае нам важно показать общую структуру критерия: делается определенное предположение, которое затем проверяется посредством некоторого текста.

Формула χ^2 может модифицироваться в зависимости от условий и целей применения критерия.

Для чего важно определить принадлежность данного экспериментального распределения? В случае лингвистических переменных это имеет первостепенную важность: таким образом мы получаем характеристику лингвистической совокупности. Предположим, что некоторое лингвистическое распределение довольно точно аппроксимировалось бы нормальным законом. Это позволило бы в каком-то смысле приравнять рассматриваемую совокупность всем другим известным физическим совокупностям, подчиняющимся тому же за-

кону. В частности, график нормального распределения носит название «гауссова функция ошибок», основанное на том, что по Гауссу плотность вероятности для случайных ошибок астрономических наблюдений выражается формулой плотности вероятности нормального распределения.

Таким образом, наблюденное распределение лингвистической переменной входило бы в класс распределений, характеризующих такие совокупности, как ошибки наблюдений, погрешности измерений, т. е. совершенно не зависело бы от направленного выбора. Какая из лингвистических совокупностей является образованной по чисто случайным причинам? Ответ на этот вопрос сыграл бы большую роль в характеристике языковой деятельности человека.

§ 2. Г. Хердан и современное состояние фонологической статистики

Первая работа Г. Хердана «Язык как выбор и случайность» вышла в 1956 г., а его последняя работа «Современная теория языка как выбора и случайности» — в 1966 г. За эти десять лет Г. Хердан выпустил еще три книги, каждая из которых развивает и продолжает идеи, высказанные в предыдущих. Знаменательно, что в своей последней книге автор возвращается к первоначальному изложению как в смысле заглавия, так и содержательно. Он обращается к некоторым идеям, которые были намечены в первой книге, и отказывается от части новшеств, введенных им в промежутке.

Г. Хердана интересуют наиболее общие теоретические вопросы применения квантитативных методов к языкознанию. Он следующим образом характеризует особенности своего подхода: «В то время, как среди более молодых лингвистов наблюдается все большая тенденция к специализации, я всегда стремился постичь и обрисовать то, что покойный Ф. Д. Рузвельт называл «полной картиной» предмета, и меня часто приводила в отчаяние очевидная неспособность некоторых лингвистов рассматривать предмет математической лингвистики как единое целое» [2]. В этом смысле Г. Хердан идет по пути Ципфа в стремлении построить целостное здание лингвистики, объединенной вокруг единого принципа. Для Хердана математическая лингвистика — это не механический конгломерат всего, что сделано в лингвистике при помощи математики, а особая наука, базирующаяся на единой теории. С одной стороны, Хердан в центре современных структуральных устремлений, а с другой, он превосходно владеет весьма изощренным математическим аппаратом, который он стремится осмыслить лингвистически.

Концепции Г. Хердана развивались в направлении от применения к языку аппарата классической математической статистики к выра-

[2] G. H e r d a n. Quantitative Linguistics. London, 1964, стр. IX.

ботке нового аппарата, предназначенного специально для исследования лингвистических совокупностей. При этом три основных направления исследования все время прослеживаются в работах Хердана:

1. Идея статистической структуры языка как отражения основополагающей языковой дуальности.

2. Идея стабильности языковых частот.

3. Стремление найти точную форму статистического распределения лингвистических переменных.

Рассмотрим каждое из этих направлений в отдельности.

1. Первое направление исследований Г. Хердана связано с особым складом его научного мышления: от общих принципов к иллюстрации их примерами. Здесь Г. Хердан все время демонстрирует стремление оперировать общефилософскими категориями. Он рассматривает статистику как глубинную философию языка, а квантитативная лингвистика является раскрытием этой философии. Вот как сам автор определяет свою задачу в последней книге:

«Филологическая статистика (у автора literary statistics. Мы переводим это как «филологическая статистика» для обозначения всего комплекса статистических исследований языка, включая вопросы стиля и т. п. Сам Хердан употребляет термин «лингвистическая статистика» только при исследовании фонологического и морфологического уровней.— *Д. С.*) должна рассматриваться как отрасль лингвистики, а не как еще один способ применения обычной статистики.

Существуют самые противоположные концепции относительно роли статистики в изучении языка. Согласно одной точки зрения, сформулированной Болдрини («Le Statistiche letterarie e i fonemi elementari nella poesia». Milano, 1948), язык рассматривается всего лишь как еще один объект для применения количественных методов, независимо от того, представляют результаты такого количественного исследования интерес для лингвиста или нет. Если я правильно понял аргументы Болдрини, он считает, что статистические результаты достаточно интересны сами по себе, хотя этот интерес не обязательно должен быть лингвистическим. На другом полюсе точка зрения А. С. Росса (A. S. Ross, op. cit.) о том, что теория вероятности и статистика должны быть инструментами или, как мы говорим сегодня, математическими моделями для проверки и уточнения любых заключений в лингвистике, которые содержат количественные данные. Согласно этой точке зрения, математические методы являются лишь вспомогательными инструментами в лингвистическом исследовании.

Одним из преимуществ систематического изложения предмета является то, что оно позволяет нам исправлять ошибочные взгляды, подобные вышеприведенным, которые возникли в то время, когда еще не было полной картины предмета. Концепция филологической статистики как квантитативного изложения теории

структурной лингвистики де Соссюра и как научного построения, дающего материал для понимания языка как системы кодирования, показывает различные попытки применения статистических методов к языковому материалу в их подлинной перспективе...

Если бы филологическая статистика была всего лишь одним применением научного метода, то она не обязательно должна была бы обладать внутренней лингвистической ценностью; как особая отрасль лингвистики филологическая статистика непременно должна представлять интерес для лингвиста, ибо в противном случае ею не будут заниматься...

Проверка лингвистических выводов, основанных на многочисленных наблюдениях, методом статистических тестов несомненно находит применение в нашей науке, но эта проверка еще не является самой наукой. Филологическая статистика не только не исчерпывается применением критериев значимости, но сама проверка по этим критериям отнюдь не является основной функцией филологической статистики. Проверка гипотез при помощи статистических методов является в настоящее время общепринятой процедурой и вполне естественно распространить ее и на лингвистические гипотезы, не превращая эти тесты одновременно в своего рода верховного арбитра. Несмотря на то, что в принципе статистические тесты вполне приемлемы, они не всегда уместны для решения конкретных лингвистических задач. Поэтому возникает необходимость в более совершенной статистике, чьи данные явились бы более убедительным аргументом, чем показания обычных критериев значимости... Филологическая статистика как квантитативная философия языка должна быть применима ко всем отраслям языкознания.

По нашему мнению, филологическая статистика — это структурная лингвистика, поднятая на уровень квантитативной науки или квантитативной философии» [3].

Здесь мы можем проследить стремление Хердана эмансипировать квантитативную лингвистику от чисто прикладных задач, его убеждение в том, что статистика и является подлинной лингвистикой, отражающей глубинные антиномии языка.

Продолжим наш обзор основных методологических положений Г. Хердана.

Глубинная система квантитативной лингвистики определяется, по Хердану, тем, что язык пронизывает основополагающая философская дуальность — независимость звука и значения (означающего и означаемого, по де Соссюру):

«Самое важное с точки зрения математической лингвистики — это независимость звука от значения, подчеркивавшаяся де Соссюром... Я называю это Аксиомой 1, поскольку она имеет фундамен-

---

[3] G. Herdan. The Advanced Theory of Language as Choice and Chance, стр. 9—11.

тальное значение для применения статистики на уровне фонологии. Если бы это не было так, то одно и то же понятие не могло бы выражаться на разных языках и различными словами. Если эта аксиома верна, то явно неслучайная последовательность слов в литературном тексте или в высказывании устного языка должна дать случайную выборку звуков, фонем, а также букв, поскольку критерий, с помощью которого мы производим выборку, а именно слова, расположенные согласно их значению, не коррелирует с критерием самой выборки, а именно индивидуальными звуками языка и буквами алфавита» [4].

Из вышеприведенной цитаты мы видим, что между общими положениями, которые Г. Хердан кладет в основу своей теории, и конкретной языковой реальностью существует некоторый барьер. Дело в том, что в принципе положение о дуальности звучания и значения абсолютно верно, однако проблема не столь однозначно проста, как это рисует Хердан, и переход от общего положения к его статистической интерпретации несколько немотивирован.

Из дуальности «означаемое—означающее», согласно Хердану, следует, что любой текст априори должен дать случайную выборку относительно частот фонем. Если не углубляться в проблему немотивированности знака [5], то даже на уровне обсуждения, предлагаемом Г. Херданом, можно заметить, что независимость значения от звучания постулируется де Соссюром для отдельного слова, а не для текста. В самом деле, очевидно, что между понятием, выражаемым цепочкой русских фонем [pr'ijóm]— «прием», и самой этой цепочкой нет причинно-следственной связи. Однако если взять текст, который можно реально услышать по радио: «Прием, прием. Прием, прием и т. д.», причем эти слова могут повторяться довольно долго, то столь же очевидно, что этот текст никак не может считаться случайной выборкой относительно частот фонем, хотя отдельное слово и обладает абсолютной независимостью звучания от значения. Могут возразить, что этот пример специально подобран, однако ниже мы продемонстрируем, что, например, поэтические тексты также устроены довольно неслучайно в смысле аранжировки фонем. Где граница между такого рода текстами и текстами, которые действительно могут представлять собою случайную выборку фонем? Каков критерий определения этой границы?

Хердана эти вопросы совершенно не волнуют. И здесь, как нам представляется, проявляется основной порок его метода — полная неозабоченность исследованием реального материала для проверки основных положений. Отсюда постулирование аксиом, которые, возможно, и не являются аксиомами, выдвижение большого коли-

---

[4] G. H e r d a n. Quantitative Linguistics, стр. IX.

[5] См., в частности, нашу заметку «Немотивированность знака» («Тезисы докладов во Второй летней школе по вторичным моделирующим системам». Тарту, 1966, стр. 12—14).

чества общих принципов и законов коммуникации, которые отчасти повисают в воздухе [6].

Однако посмотрим, каково дальнейшее распространение принципа дуальности у Хердана. Совершенно естественно, после оппозиции «означающее — означающее» анализируется оппозиция «речь—язык». Вот что пишет сам Г. Хердан:

«Моя теория о том, что математическая и особенно статистическая лингвистика — это не что иное, как квантификация соссюровской теории языка, получила желанное подтверждение: квантитативная лингвистика представляется теперь тем, что де Соссюр имел в виду под термином langage в отличие как от langue, так и от parole.

Огромная важность этого трехчленного деления признавалась Блумфильдом («Modern Language Journal», 1924, № 8): «В любое данное время (синхронно) язык общества следует рассматривать как систему сигналов... Эта строгая система, предмет дескриптивной лингвистики, как мы говорим,— la langue (язык). Но le langage (человеческая речь) включает в себя нечто большее, поскольку индивидуумы, составляющие общество, не могут следовать системе с абсолютным единообразием. Реальная речь, высказывание (la parole) изменяется не только в тех аспектах, которые не фиксируются системой,— ср., например, точный фонетический характер каждого звука; она изменяется и в том, что касается самой системы: различные говорящие могут в любой данный момент нарушить почти любой признак системы».

То, как Блумфильд формулирует свое положение, а именно, что le langage (человеческая речь) содержит нечто в дополнение к la langue, удивительным образом совпадает с моей концепцией отношения langue — parole как отношения между статистической совокупностью и выборкой. Стабильность относительных частот, прикрепленных к различным членам данного рода лингвистических форм, неизбежно приводит к заключению, что la langue включает в себя не только языковые элементы как лексические формы, но эти элементы плюс соответствующие вероятности их появления. Это — то, что я назвал статистической интерпретацией соссюровской дихотомии. Основной закон коммуникации сводится, таким образом, к утверждению, что язык — это обобщающий термин для обозначения языковых элементов (фонем, слов, метрических форм)

[6] Ср. следующее категорическое высказывание в книге «Quantitative Linguistics» (стр. XIV): «Различные комбинации фонем, принадлежащие данной цепочке (морфеме), не имеют никакого отношения друг к другу кроме того, что они представляют собою случайные вариации (комбинации) одних и тех же основных единиц». Это высказывание приводится в поддержку мысли автора о том, что тексты представляют собой случайные выборки фонем. Однако достаточно вспомнить такие факты, как отношения сингармонизма, ассимилятивные и диссимилятивные процессы, нейтрализация и т. д. и т. п., чтобы увидеть, что приведенный выше постулат нуждается по крайней мере в уточнении.

вместе с вероятностями их появления (G. Herdan, Language as Choice and Chance, p. 79—81).

Теперь оказывается, что «la langue» плюс вероятности различных форм, которые входят в «langue» — это как раз то, что Блумфильд имел в виду под «langage». «Нечто в дополнение»— это вероятности, появляющиеся в человеческой речи при употреблении langue» [7].

Как мы видим, здесь Хердан связывает дихотомию звучания и значения с дихотомией «язык—речь», причем последнее отношение характеризуется стабильностью относительных частот. К этому пункту мы вернемся ниже, а здесь укажем, что сама проблема места статистического аспекта языка в соотношении «язык—речь» является совершенно неисследованной. Это и понятно, поскольку лингвистическая статистика еще не накопила достаточно данных, чтобы можно было более или менее обоснованно рассуждать об организации статистического аспекта языка. Поэтому стремление Хердана раз и навсегда решить этот крайне сложный вопрос не может вызвать сочувствия. С другой стороны, представляется, что языковая статистика действительно может иметь отношение к «речи вообще». Этот аспект очень важен и до настоящего времени почти не исследован. Сюда входят не только вероятностные аспекты, но и такие моменты, как организация высказываний, больших, чем предложение [8], исследование реальных значений грамматических категорий и т. п.

В своей последней книге Г. Хердан в духе своей первой монографии развивает идеи статистической интерпретации противопоставления «язык—речь». Приведем цитаты из введения к этой книге:

«Совокупность элементов, употребляемых данным языковым сообществом, представляет формы или варианты, в виде которых эти элементы могут выступать. Данные формы совместно с соответствующими вероятностями появления составляют распределение плотности вероятности определенной лингвистической переменной. Конечно, эти вероятности не известны нам заранее, и мы можем их оценить лишь по относительным частотам, наблюдаемым в выборках индивидуальной речи. Но таково положение со всеми распределениями плотности вероятности, основывающимися на наблюденных данных, и лингвистические распределения не являются исключением. В то время, как индивидуальная встречаемость в речи должна трактоваться лишь как выборка из всей совокупности,

---

[7] G. Herdan. Quantitative Linguistics, стр. 3—4.

[8] См., в частности: Е. В. Падучева. О структуре абзаца. «Труды по знаковым системам, II». Тарту, 1966; И. П. Севбо. Об изучении структуры связного текста. «Лингвистические исследования по общей и славянской типологии». М., 1966, а также нашу статью «О связи семантики текста с его формальной структурой» («Poetics. Poetyka. Поэтика II». Warszawa, 1966).

накопление очень большого числа таких выборок позволяет получить оценку вероятностей в генеральной совокупности.

...В той мере, в какой индивидууму приходится употреблять слова своего языка, и у него нет свободного выбора в употреблении элементарных звуков, из которых эти слова состоят, его языковое поведение находится ниже индивидуального контроля. Соссюровская аксиома независимости звучания и значения имеет своим следствием постоянство относительной частоты фонем независимо от содержания высказывания или текста. Это приводит к концепции определенных вероятностей, приписываемых различным фонемам, и к понятию случая как лингвистического фактора» [9].

Как явствует из приведенных высказываний Г. Хердана, постулат о стабильности частот языковых форм (в частности, фонем) является для него настолько теоретически очевидным, что не нуждается в экспериментальной проверке. Некритически воспринятые общие положения де Соссюра экстраполируются на область, по-видимому, принципиально отличную от языка в соссюровском понимании.

В то же время ценность рассуждений Г.Хердана в том, что они привлекают внимание к постулату о стабильности частоты, ставят его в центр лингвистической проблематики.

Однако было бы несправедливым не отметить, что за истекшие десять лет давление лингвистической реальности привело к тому, что Г. Хердан вынужден был изменить некоторые основы своей теории:

«В моей книге «Language as Choice and Chance» были заложены основы действительно плодотворного применения статистики к языку вследствие того, что я интерпретировал дихотомию «язык — речь» как дихотомию между «статистической совокупностью» и «выборкой».

Однако «течение истинной любви не знает постоянства», и в этом смысле не являются исключением интимные отношения между статистической теорией и лингвистическим наблюдением. Короче говоря, постулат о соответствии дихотомии «речь — язык» отношению между статистической генеральной совокупностью и случайной выборкой оказывается справедливым только для мельчайших элементов языка — фонем и букв, к которым, однако, соссюровская дихотомия не относилась. С другой стороны, в случае дихотомии словника связного текста и словаря, которую и имел в виду де Соссюр, необходима определенная модификация, если мы хотим, чтобы было выполнено условие случайного выбора из статистической совокупности.

---

[9] G. H e r d a n. The Advanced Theory of Language as Choice and Chance. Introduction.

Связные тексты обязательно имеют дело с определенным конкретным содержанием, поэтому их нельзя рассматривать как случайные выборки из словаря языка» [10].

Оставим в стороне утверждение о том, что де Соссюр имел в виду именно дихотомию между словарем и текстом, говоря о различии языка и речи. Это настолько явная натяжка, что нет смысла полемизировать. Существенно другое: автор вынужден поставить под сомнение основную аксиому своей концепции, а именно — что тексты представляют собой случайную выборку из всего словаря языка. Правда, здесь эта оговорка делается лишь в отношении словаря, но это позволяет взглянуть на проблему еще раз и проверить основные положения. Эта проверка приводит Хердана к формулированию так называемой «новой статистики» для словаря. Представляется, что и для фонемного уровня постулат о связном тексте как о случайной выборке фонем подлежит проверке [11]. Г. Хердан сделал уступку в отношении словаря, в отношении же фонем и букв он остается на прежних позициях, поэтому наша задача — подвергнуть критике концепцию Хердана и в этом плане. Приходится, к сожалению, отметить, что в объемистых книгах Хердана главное для самого ученого — это защита достаточно сомнительных общих положений. Сама постановка вопроса о соотношении «язык — речь» и статистическом аспекте языка в настоящее время, когда еще неясны учетные единицы и уровни абстракции лингвостатистики, просто некорректна.

2. Если положение о статистической совокупности как репрезентанте понятия «язык» и о статистической выборке как репрезентанте понятия «речь» самому Г. Хердану пришлось подвергнуть частичной ревизии, то вторая кардинальная аксиома квантитативной лингвистики — стабильность лингвистических частот — остается неизменной на протяжении всех десяти лет. Вот цитата из последней книги:

«Замечательный факт стабильности относительных частот лингвистических символов оказывается общей характеристикой языковых форм. Мы встречаемся с ним в распределении относительной частоты фонем, букв, длины слова в терминах количества букв и слогов, грамматических форм и латинских и греческих гекзаметров согласно их метрической структуре.

---

[10] G. Herdan. The Advanced Theory of Language as Choice and Chance. Preface.

[11] Ср. следующее высказывание Р. Абернати в его докладе «Some Theories of Slavic Linguistic Evolution» на V Международном съезде славистов в Софии: «Формула, которую дает Г. Хердан в своей книге «Type-Token Mathematics» и про которую он утверждает, что она предсказывает частоту фонем по их «словарным частотам», на самом деле не предсказывает абсолютно ничего». («American Contributions to the Fifth International Congress of Slavists». The Hague, 1963, стр. 17).

Было бы ненаучным остановиться на этом и удовлетвориться простой констатацией факта стабильности для каждого из этих распределений в отдельности.

Поскольку совершенно очевидно, что во всех этих распределениях присутствует некоторый общий элемент, а именно, что все они — часть механизма языковой коммуникации, представляется разумным найти общие выражения этой регулярности прежде чем искать ее объяснения. Формулировка, обобщающая эмпирические наблюдения, рассмотренные выше и встречающиеся во многих других рядах лингвистических форм, будет звучать следующим образом: «Пропорциональное содержание лингвистических форм, принадлежащих к одному языковому уровню или к одной стадии языкового кодирования — фонологической, грамматической, метрической,— остается в общем постоянным для данного языка в данный период его развития и для достаточно большого числа однородных наблюдений» [12].

Разумеется, выделение метрического уровня как отдельного от фонологического может вызвать справедливые нарекания. Видимо, целесообразнее рассматривать метрику как часть фонологического уровня. Далее, частотное распределение фонем и длины слов — суть явления разного порядка. В первом случае мы имеем дело с качественной случайной величиной, а во втором — с количественной. Поэтому «основной закон коммуникации», выведенный Херданом, относится не к однородным наблюдениям, о которых в нем говорится, а к наблюдениям над принципиально различными объектами.

Однако основную претензию следует предъявить к фактическому обоснованию постулата о стабильности языковых частот. Во всех пяти книгах Г. Хердан использует один и тот же материал: данные о частотности фонем в итальянской поэзии (согласно Болдрини), сравнение частотности чешских фонем в нескольких небольших (приблизительно по 700 фонем) выборках из Чапека и данные о частотности английских фонем Дьюи [13]. Помимо этого были использованы материалы о распределении длины слов в терминах количества букв и слогов, о частотности грамматических форм (данные подсчетов по «Капитанской дочке») и о распределении латинских и греческих гекзаметров. Все эти материалы не могут быть признаны ни достаточными, ни тем более исчерпывающими. В частности, для Хердана, по-видимому, несущественно, что в выборках из Чапека не обеспечивается даже однократная встречаемость всех фонем. Он суммирует все малые частоты, произвольно уменьшая количество сравниваемых единиц, продолжая при этом называть

---

[12] G. Herdan. The Advanced Theory of Language as Choice and Chance, стр. 5 и след.

[13] G. Dewey. Relative frequency of English Speech Sounds. Cambridge, 1925.

их фонемами (например, если оказывается, что частоты чешских фонем $\bar{o}$, g и $\dot{d}$ слишком малы, то они суммируются, таким образом получается новая частотность, которая на фонологическом уровне ничему не соответствует, так как $\bar{o}$, g и $\dot{d}$ не образуют фонологически значимого класса). Разумеется, в статистике процедура уменьшения количества сравниваемых разрядов является вполне общепринятой. Однако разряды, получающиеся после уменьшения, должны семантически соответствовать первоначальным разрядам.

Столь же проблематичны и данные о стабильности частот грамматических форм. Они вызывают больше всего сомнений. Выборки, послужившие основой для подсчетов, весьма невелики и непредставительны. Поэтому выводы, которые делает Г. Хердан, являются скорее повторением исходного постулата, чем его доказательством.

С другой стороны, за последнее время появились работы, в которых на обширном материале доказывается прямо противоположное тому, что утверждал Г. Хердан, а именно, что грамматические элементы обладают различной частотой в различных типах текстов (функциональных стилях), так что о стабильности общеязыковой частоты не может быть и речи. Мы имеем в виду замечательные работы Г. А. Лесскиса [14] о грамматической структуре предложений в различных видах русской прозы XIX в. Г. А. Лесскис показал, что пропорциональное содержание различных классов слов (в том числе и служебных слов) в простых и сложных предложениях соответственно художественной и научной прозы является существенно различным. Данные Г. А. Лесскиса, полученные на громадных массивах текста, являются прекрасным доказательством того, что проблема стабильности частот грамматических классов является гораздо более сложной, чем ее представлял себе Г. Хердан (и ранее П. Гиро). Во всяком случае, ни о какой стабильности относительных частот грамматических классов для общеязыковой совокупности после работ Г. А. Лесскиса уже нельзя говорить.

Наконец, в самое последнее время появилась небольшая заметка Н. П. Савицкого [15], в которой проблема стабильности относительно языковых частот ставится непосредственно в связи с работами Г. Хердана: «Хердан (G. Herdan. Language as Chance and Choice. Groningen, 1956, стр. 66—96) и Черри (B. C. Cherry. On Human Communication. N. Y. 1957, стр. 117) показали, что возможность описания естественного языка в понятиях теории инфор-

---

[14] См., в частности: Г. А. Л е с с к и с. О зависимости между размером предложения и его структурой в разных видах текста.— ВЯ, 1964, № 3, стр. 99—123; О н ж е. К вопросу о грамматических различиях научной и художественной прозы.— Сб. «Σημειωτική», стр. 76—83; О н ж е. Два способа описания внеязыковых ситуаций. «Лингвистические исследования по общей и славянской типологии». М., 1966, стр. 32—50.

[15] Н. П. С а в и ц к и й. Об устойчивости относительных частот лингвистических элементов. «Ceskoslovenská rusistika», 1966, № 4, стр. 214—217.

мации существенным образом зависит от устойчивости относительных частот лингвистических элементов (фонем, слов, частей речи и др.), т. е. от того, в какой мере относительные частоты этих элементов, установленные на разных текстах или на разных выборках из одного и того же текста, совпадают между собой. Проблема устойчивости частот лингвистических элементов или, иначе говоря, проблема статистической однородности текстов относительно этих частот представляет собой, таким образом, важный аспект более общей проблемы возможностей математического моделирования языка.

Хердан указал также на роль грамматики как фактора, регулирующего относительные частоты лингвистических элементов. Однако по вопросу о том, в какой же мере эти частоты в разных текстах действительно совпадают, он ограничился довольно туманными высказываниями.

При сравнении двух и более серий относительных частот элементов некоторого рода возможны в основном три случая:

1. Различия между сериями столь незначительны, что их можно считать случайными отклонениями от одного основного распределения. В этом случае мы можем полученные серии считать выборками из некоей основной совокупности, обладающей определенными вероятностями появления элементов рассматриваемых типов.

2. Различиями между сериями нельзя пренебречь, эти различия статистически значимы, но в то же время между распределениями частот в рассматриваемых выборках имеется все же некоторое сходство (между ними имеется значимая корреляционная связь).

3. Различия между сериями столь велики, что между ними невозможно найти никакого неслучайного сходства» [16].

Здесь ставится, по существу, та же задача исследования однородности текстов, которая ранее был сформулирована Р. М. Фрумкиной [17] и подробно рассматривалась в нашей статье [18].

И. П. Савицкий использует данные Н. Я. Меца [19] и показывает, что частотность функциональных классов имени существительного в латышском языке различна даже в пределах выборки из текстов с максимальным уровнем стилистического сходства. Применяется критерий χ^2 по формуле, приведенной выше (теоретические вероятности вычисляются в предположении, что распределение классов существительных во всех выборках одинаково). Делается предва-

---

[16] Н. П. Савицкий. Указ. соч., стр. 214.

[17] Р. М. Фрумкина. О законах распределения слов и классов слов. «Структурно-типологические исследования». М., 1962, стр. 124—133.

[18] Д. М. Сегал. Статистическая однородность текста на фонологическом уровне в польском языке. «Структурная типология языков». М., 1966, стр. 26—44.

[19] Н. Я. Мец. Некоторые статистические характеристики имен существительных и глаголов в латышском языке. «Статистико-комбинаторное моделирование языков». М.— Л., 1965, стр. 356—365.

рительный вывод о том, что грамматика языка накладывает большие ограничения на статистическую структуру текстов, но эти ограничения все же недостаточно сильны для того, чтобы тексты образовали статистически однородную совокупность.

Таким образом оказывается, что один из основных постулатов теории квантитативной лингвистики Г. Хердана — постулат об обязательной стабильности лингвистических частот — в некоторых случаях оказывается неверным. Вспомним, что именно этот постулат в явном или неявном виде лежал в основе всех рассмотренных нами выше построений лингвистической статистики. Проверка постулата о стабильности лингвистических частот представляется нам, в силу его важности для лингвистической статистики, ее основной задачей. Без установления на самом различном материале и при помощи разнообразных статистических процедур того, насколько стабильны (или нестабильны) частоты лингвистических элементов, дальнейшее развитие лингвистической статистики как лингвистической науки невозможно. Постулат стабильности частот, помимо прочего, является весьма конкретной статистической пропозицией (в отличие от «принципа дуальности», который, кстати, базируется на стабильности частот), поэтому этот постулат сравнительно легко проверить, — это не связано с теоретическими контроверзами и имеет большую практическую важность.

Итак, проблема стабильности лингвистических частот лежит в самом центре лингвистической статистики. Недоговоренности и туманные места у Хердана лишь стимулируют исследование этой важной задачи.

Говоря о стабильности лингвистических частот, Г. Хердан пишет, что наилучшим инструментом для проверки стабильности является критерий χ^2. Этот критерий Г. Хердан считает основным аппаратом фонологической статистики. При этом Хердан сознает, что возможны и такие случаи, когда применение критерия χ^2 обнаружит нестабильность частот (хотя в приводимых им примерах этого не происходит). Он пишет по этому поводу следующее:

«В терминах теории статистики мы описываем подобную неоднородность (т. е. нестабильность частот.— *Д. С.*) как невыполнение в подобных совокупностях условий случайного отбора. Как показал английский статистик Лексис (Lexis), в совокупностях, относящихся к явлениям общественной жизни человека, условия случайной выборки почти никогда не выполняются так, как это имеет место при бросании костей или в других играх, основанных на случайности. При этом выборка становится тем менее случайной, чем шире поле наблюдения. Распределения лингвистических элементов, которые Блумфильд назвал мельчайшими единицами общественной жизни, в этом отношении не являются исключением» [20].

[20] G. Herdan. The Advanced Theory of Language as Choice and Chance, стр. 43.

Несмотря на то, что частоты могут оказаться нестабильными в смысле критерия χ^2, для Хердана совершенно очевидно, что в принципе стабильность частот фонем является настолько фундаментальным фактом, что несогласие наблюденных частот с гипотезой стабильности объясняется лишь несовершенством критерия: «Без достаточно стабильных рядов относительных частот лингвистических символов или форм невозможно никакое предсказание или, точнее говоря, правильная отгадка пропущенных частей сообщения. В этом заключается связь между стабильностью относительных частот языковых элементов и информацией. Тот факт, что эти распределения не всегда стабильны с точки зрения теории статистической случайной выборки, не мешает тому, чтобы они были достаточно устойчивы для обеспечения коммуникации» [21].

Здесь Г. Хердан затрагивает наиболее важный аспект проблемы стабильности языковых частот. С одной стороны, исследования Г. А. Лесскиса и других показали, что частотность грамматических элементов сильно зависит от содержания текста и, следовательно, не является общеязыковой; с другой стороны, существует интуитивное представление о том, что одни лингвистические элементы (фонемы, грамматические элементы и т. п.) встречаются чаще, чем другие. Возможно ли примирить обе точки зрения? Г. Хердан подходит к выводу, что следует модифицировать сам критерий установления стабильности, чтобы он смог эту истинную стабильность выявить. Данный вывод, правда, не формулируется в явном виде, но вся логика рассуждений Г. Хердана свидетельствует об этом.

Тот факт, что этот вывод не формулируется в явном виде, говорит об известной непоследовательности Г. Хердана. С одной стороны, он настолько глубоко уверен в постоянстве языковых частот, что готов даже отказаться от статистического аппарата (таким образом, для Г. Хердана статистические приемы — это не самоцель), с другой стороны, неумение сделать шаг от общих деклараций к конкретным исследованиям мешает ему видеть пути модификации своего тезиса: Хердан не пытается проанализировать структуру возможной нестабильности частот, не пытается отделить элементы, у которых частота более стабильна, от элементов, частота которых более подвижна. Привязанность к тезису не перерастает в стремление доказать его, тезис лишь декларируется.

Тем не менее тот факт, что Хердан видел, что с помощью существующих критериев стабильность частот удается установить не всегда, и, несмотря на это, продолжал считать стабильность частот аксиомой, позволяет пойти дальше, позволяет попытаться установить, что же в речевой деятельности человека действительно стабильно, а что подвержено флуктуации. Заметим в заключение этого раздела, что Хердан преувеличивает роль стабильности час-

[21] Там же.

тоты в процессе коммуникации. Если бы каждый элемент обладал настолько стабильной частотой, что возможность вариации была бы сведена к минимуму, то существенно уменьшилась бы гибкость языкового кода, характеристика, особенно важная для языка как средства коммуникации.

3. Как мы пытались продемонстрировать это в пунктах 1 и 2, Г. Хердан, стремясь оставаться в рамках общетеоретической концепции, все же видит те пункты, в которых эта концепция наиболее слаба, и в которых одновременно возможен наибольший теоретический прогресс при некотором новом подходе.

Наиболее ярко эта способность изменять теорию при столкновении с трудностями проявляется там, где Г. Хердан трактует вопросы теории лингвистических распределений. Мы специально остановимся на этом, поскольку данный раздел представляет собой особый общестатистический интерес, хотя наиболее интересные мысли Г. Хердана относятся к словарной статистике, выходящей за рамки настоящей работы.

В своей первой книге «Язык как выбор и случайность» Г. Хердан уделил сравнительно немного места вопросу о характере лингвистического частотного распределения. Собственно, он пишет об этом лишь мимоходом там, где говорит о выработке статистического критерия для оценки богатства словаря. Поскольку этот критерий исходит из того, что отношение логарифма количества слов в словнике данного текста к логарифму общего количества слов во всем тексте $\frac{\log V}{\log N}$ является постоянным для данного литературного текста, Хердан делает вывод, что частотное распределение лингвистических объектов описывается модификацией нормального закона, так называемым логарифмическим нормальным законом, в котором вместо значений случайной величины выступают логарифмы этих значений $\log \xi$.

Больше ничего по этому вопросу в первой книге не содержится.

Во второй книге «Лингвистическая математика» Г. Хердан посвящает вопросу о законе распределения лингвистических переменных целую главу.

Г. Хердан пишет, что было сделано несколько предположений относительно характера закона, по которому распределяются лингвистические случайные величины. Юл предположил, что частотные распределения слов подчиняются закону Пуассона. Г. Хердан пишет, что предположение о распределении лингвистических случайных величин по закону Пуассона надо отвергнуть по двум причинам: 1) чисто случайный характер пуассоновского распределения не соответствует ситуации в языке; 2) формула пуассоновского распределения влечет за собой слишком трудоемкие вычисления, что невыгодно.

Г. Хердан подробно анализирует так называемый «закон Ципфа» и приходит к выводу, что он неприемлем в качестве общего

закона лингвистических распределений. Далее Г. Хердан снова выдвигает предположение о том, что логарифмический нормальный закон лучше всего описывает распределение лингвистических объектов. Приходится отметить, что это предположение является лишь гипотезой, но выдвигается как нечто вполне доказанное.

В нашей литературе за последние несколько лет появились работы, посвященные экспериментальной проверке предположений о том, что распределение лингвистических объектов подчиняется нормальному закону, или закону Пуассона. Это, в частности, работы Р. М. Фрумкиной «О законах распределения слов и классов слов» [22] и Л. А. Турыгиной и М. Н. Боркуна «Статистические методы исследования частотного распределения лингвистических единиц» [23].

Р. М. Фрумкина проверяет гипотезу о распределении Пуассона для частот слов *день*, *новый*, *другой*, *знать*, *без*, *при*, *до*, *над*, *между*, *очень*, *здесь*, *или* в «Словаре языка Пушкина».

Применение критерия χ^2 для проверки близости наблюденных частот к частотам, вычисленным в предположении пуассоновского распределения, показало, что для слов *без*, *другой*, *или*, *над* и *между* гипотеза не противоречит данным наблюдения, в то время как для слов *день*, *знать*, *очень*, *при*, *новый* эта гипотеза должна быть отвергнута. Все рассмотренные слова принадлежат к словам с довольно высокой частотой в «Словаре языка Пушкина», поэтому трудно найти статистические обоснования тому, что гипотеза для одних слов не отвергается, а для других отвергается. Равным образом однозначного семантического или грамматического различия между этими словами не наблюдается. Р. М. Фрумкина пишет, что «преждевременно было бы пытаться... предлагать какие-либо объяснения тому обстоятельству, что проверяемые статистические гипотезы в ряде случаев не подтверждаются» [24]. Далее, в той же работе проверяется гипотеза о нормальном распределении, но уже не для отдельных слов, а для классов слов: существительных, прилагательных и глаголов. Эта гипотеза хорошо подтверждается для глаголов и хуже для существительных, для прилагательных ее следует отвергнуть вовсе. Р. М. Фрумкина пишет: «Характер расхождений между наблюденными частотами и вычисленными в предположении нормального распределения наводит на мысль о том, что, быть может, распределение классов слов ближе к логарифмически нормальному» [25].

Таким образом, Р. М. Фрумкина делает то же предположение о частотном распределении слов, что и Г. Хердан.

---

[22] «Структурно-типологические исследования». М., 1962, стр. 124—133.
[23] «Энтропия языка и статистика речи». Минск, 1966.
[24] Р. М. Фрумкина. О законах распределения слов и классов слов, стр. 132.
[25] Там же.

Данные Р. М. Фрумкиной отчасти совпадают с данными Л. А. Турыгиной и М. Н. Боркуна. Указанные авторы исходят из предположения о том, что употребление того или иного слова в тексте есть случайное событие, независимое от употребления других слов и имеющее постоянную вероятность p.

В таком случае вероятность p(N, m) появления фиксированной словоформы m раз при прочтении в тексте N словоформ определяется формулой биномиального распределения: $p(N,m) = C_N^m p^m q^{N-m}$, где $q = 1 - p$. При достаточно большой выборке N и m, значительно превышающем 1, биномиальное распределение можно аппроксимировать распределением Гаусса. Берутся две выборки из однородных текстов объемом в N_1 и N_2 словоформ. Авторов интересует вопрос о вероятностях употребления некоторой словоформы P_1 и P_2 в этих двух выборках. Для исследования этого вопроса удобно перейти к распределению вероятности относительно частот в этих выборках $\frac{M_1}{N_1} = x_1$ и $\frac{M_2}{N_2} = x_2$, колебания которых, исходя из сделанного допущения, будут описываться распределением Гаусса с математическим ожиданием $M(x) = M\left(\frac{m}{N}\right) = p$ и дисперсией $\sigma^2(M) = \sigma^2\left(\frac{M}{N}\right) = \frac{M}{N}$.

Далее авторы берут разность относительных частот $x_1 - x_2 = y$, где y — случайная величина, если x_1 и x_2 — случайные величины. Тогда распределение случайной величины y также описывается нормальным законом. Далее вводится случайная величина z (нормированная разность относительных частот), равная $\frac{y}{\sqrt{\sigma_1 + \sigma_2}}$, приводящая к универсальной форме нормального распределения. Распределение z для слов первых 208 рангов сравнивается с теоретическим распределением при помощи критерия χ^2. Следует отметить, что здесь эксперимент выполняется на огромном материале (208 слов!). Применение нормированной случайной величины z обеспечивает большую устойчивость и представительность распределения. Таким образом, рассматриваемый эксперимент является наиболее авторитетной проверкой гипотезы о нормальном распределении частот отдельных слов.

Получившееся значение χ^2 оказывается слишком большим, и гипотеза о нормальном распределении частот словоформ должна быть отвергнута. Затем авторы рашают проверить гипотезу о нормальном распределении не на отдельных словах, а на частях речи. Результаты применения критерия χ^2 показывают, что эта гипотеза не отвергается для всех классов, кроме существительных и местоимений. Таким образом, и здесь гипотеза о нормальном распределении проходит не для всех классов слов (так же, как и в работе Р. М. Фрумкиной эта гипотеза не проходит для существительных).

Итак, экспериментальные работы показывают, что пока не найдено такого закона распределения лингвистических вероятностей, который подходил бы для всех языковых объектов.

Это замечает и Хердан, который в своих книгах «Количественная лингвистика» и «Исчисление языковых наблюдений» отказывается от идеи найти единый закон распределения вероятностей для всех языковых объектов и вводит понятие смешения распределений. На этом этапе Хердан считает, что фонологический уровень описывается так называемым мультиноминальным распределением (модификация упомянутого выше биноминального распределения), которое при соответствующих условиях хорошо аппроксимируется логарифмически нормальным распределением. На лексическом уровне статистическое распределение наблюдаемых частот может быть описано как смешение трех различных распределений. Эти три распределения отражают тот факт, что словарь может быть разделен на три зоны: слова с высокой частотой, слова со средней частотой и редкие слова, встречающиеся в тексте всего один — два раза. Предлагается описывать распределение вероятности частых слов биномиальным распределением, аппроксимируемым логарифмически нормальным законом (Хердан считает, что распределение частых слов аналогично распределению фонем). Слова средней частоты распределяются по сложному закону Пуассона, а редкие слова — по простому закону Пуассона.

И в этом случае предположения о характере законов распределения не проверяются, а высказываются в общем виде.

Как явствует из работы Р. М. Фрумкиной, подобные априорные утверждения, по-видимому, окажутся неподтвержденными: закон Пуассона одинаково принимается и отвергается для слов с высокой частотой.

Однако Г. Хердан не удовлетворяется тем, что распределение вероятности слов описывается как смешение трех различных распределений. В книге «Количественная лингвистика» он вновь пытается найти общую форму распределения для всего словаря в целом. Исходя из общей схемы порождения вероятностей слов, аналогичной, по мнению Г. Хердана, распределению частиц разного размера в смесях, предлагается так называемое распределение Уоринга (Waring distribution), которое объявляется основным законом статистического распределения словаря. Мы не будем здесь описывать это распределение, тем более что в дальнейшем Хердан от него отказывается. Отметим лишь, что при проверке совпадения наблюденных частот слов с частотами, вычисленными в предположении распределения Уоринга, согласие при помощи критерия χ^2 оказывается очень слабым, и Хердану приходится прибегать к объединению частот редких слов, чтобы его улучшить.

Неудача попыток найти общий закон статистического распределения всех лингвистических объектов приводит Г. Хердана к пересмотру некоторых основных положений статистической теории

применительно к словарю. Вопрос о законе распределения фонем мало интересовал его и вначале (он считал, что какая-то модификация биномиального распределения, аппроксимируемая логарифмически нормальным законом, описывает распределение фонем); в последней книге эта проблема не затрагивается совсем. Зато основной упор в своей теории квантитативной лингвистики Г. Хердан теперь делает на выработку так называемой «новой статистики», которая должна более адекватно описать поведение слов в массовых совокупностях. Статистика на фонологическом уровне и статистика на уровне словаря становятся двумя совершенно различными математическими дисциплинами с различными исходными положениями. Эти различия можно суммировать следующим образом:

А. Фонемная статистика. Здесь применяются классический аппарат статистики и основные положения теории вероятностей. Наиболее существенной теоретической предпосылкой фонемной статистики является постулат о том, что генеральная статистическая совокупность здесь бесконечна и представляет весь язык. Отдельные тексты являются конечными случайными выборками из этой бесконечной совокупности, при этом извлечение из генеральной совокупности этих выборок нисколько не меняет характер всей совокупности. Основная случайная величина — встречаемость языкового элемента в данной выборке. В данный момент времени t_i может встретиться лишь один элемент x_i. Основная характеристика этой случайной величины — ее выборочная вероятность, являющаяся репрезентантой теоретической вероятности совокупности.

Б. Словарная статистика. Здесь применяется так называемая новая статистика. Основная случайная величина меняется —это уже не встречаемость элемента в данной выборке, а встречаемость его в одной или нескольких частях текста. Характеристикой этой случайной величины является скорость повторения (repeat rate) или квадрат вероятности. Распределение вероятности того, что слово встретится в определенном количестве частей текста, предусматривает, что событие одновременно происходит в различных частях текста. Соответственно мы имеем дело не с вероятностью, а с произведением вероятностей.

Понятие бесконечной генеральной общеязыковой совокупности заменяется конечным отрезком речи, являющимся конечной статистической совокупностью, которая изменяется, когда из нее изымается конечная выборка. Таким образом, исчезает и понятие общеязыковой совокупности. Новая статистика значимым образом применима для характеристики отдельных текстов или совокупностей текстов. Случайная выборка из бесконечной совокупности заменяется случайным членением конечной совокупности.

Г. Хердан пишет, что статистическая проблема, встающая при применении новой статистики к языку,—«навешивание» номеров

мест на элементы при том, что элементов — конечное число, и они распределены по некоторым сегментам на прямой,—совпадает с проблемами статистической физики. Поэтому свою «новую статистику» Г. Хердан отождествляет со статистикой Эйнштейна—Бозе.

При современном состоянии лингвистической статистики заранее (как это хочет сделать Г. Хердан) трудно решить вопрос о том, какой тип статистики применим для того или иного уровня языка, равно как и то, какими статистическими свойствами этот уровень обладает. В принципе сейчас еще неизвестно, обладает ли язык вообще свойствами статистической совокупности. Более того, мы еще не знаем, корректно ли ставить этот вопрос. Можно делать лишь предположения, подлежащие экспериментальной проверке. Поэтому на нынешнем уровне развития лингвостатистики столь же правомерно предполагать, что весь язык (как бесконечный текст) является генеральной совокупностью, как и то, что такой совокупностью является конечный текст или словарь. Гипотез для проверки может быть много.

Равным образом заранее не известно, как обеспечить случайный характер выборки из совокупности. Поэтому одинаково равноправны различные предлагающиеся методы выбора. В принципе возможно много подходов, в частности следует экспериментально проверить, являются ли случайными выборками заведомо неслучайные тексты.

В качестве одного из подходов приведем методику, сознательно обеспечивающую максимально случайный выбор каждого элемента выборки.

В недавно опубликованной статье киевского математика Л. С. Стойковой «Об использовании выборочного метода в лингвистических исследованиях» говорится: «Случайно из ансамбля выбирают N-ое количество слов (число N называют объемом выборки). Для того, чтобы выборка содержала достаточно надежную информацию о некоторой совокупности слов, необходимо, чтобы к а ж д а я е д и н и ц а в ы б о р к и (разрядка моя.— *Д. С.*) была выбрана случайно. Это крайне важно, ибо именно случайная выборка является той математической моделью, на которую должна опираться теория статистики в связи с тем, что эта теория лежит в основе выводов, которые формулируются после эксперимента, в интересах самого исследователя приблизиться по возможности точнее к идеальным требованиям. Чем случайнее будет выборка, тем более точными будут выводы, сделанные на основе выборочных данных.

Для приведения случайной выборки из конечного ансамбля можно использовать таблицу (либо датчик) случайных чисел»[26].

---

[26] Л. С. Стойкова. До застосування вибіркового методу в лінгвістичных дослідженнях. «Статистичні та структурні лінгвістичні моделі», Київ, 1966, стр. 80—81.

Л. С. Стойкова предлагает специальный метод обеспечения случайности выборки — выбор каждого элемента с помощью таблицы случайных величин. Несомненно, что таким образом случайность будет обеспечена. Однако будет ли генеральная совокупность, откуда извлечена подобная выборка, представлять собой язык? Что скажут нам подобные выборки о статистических свойствах реальных текстов? Кроме того, подобный метод чрезвычайно трудоемок. Исходя из того, что нас в первую очередь интересуют статистические характеристики реальных текстов как массовых совокупностей, а также из очевидного равноправия методов на нынешнем этапе исследования, мы предпочитаем другой метод выбора единиц подсчета —исследование реальных текстов.

В словарной статистике, разрабатываемой Г. Херданом, имеются свои методические трудности. Прежде всего, возникает вопрос, насколько случайным должно быть членение текста. Для словарной статистики этот вопрос ставится Г. Херданом следующим образом: «Решение вопроса о том, что́ считать частью данного корпуса, является, конечно, до некоторой степени произвольным. Здесь следует руководствоваться соображениями текстового единства и математического удобства» [27]. Далее указывается, что функция практически может быть вычислена на электронно-вычислительной машине, если число частей текста не превышает 10. Таким образом, для Г. Хердана понятие случайного членения (random partitioning) представляется скорее как понятие любого членения или даже семантического (т. е. уже неслучайного) членения. Мы не будем здесь разбирать вопрос о том, какое из двух понятий — случайное членение или любое (данное) членение — адекватнее описывает статистическую структуру словаря. Укажем лишь, что, по нашему мнению, и в фоностатистике до сих пор оперировали не строгим понятием случайной выборки, а идущим от естественного языкового процесса понятием данной выборки, которая приравнивалась (может быть и незаконно) к случайной выборке.

В хердановской словарной статистике также имеется некоторый технический порог: максимальное число частей текста — 10. Это, разумеется, слишком мало, тем более, что одним из основных приложений «новой статистики» должно явиться установление авторства на основе анализа распределения словаря в различных текстах. Заведомо может встретиться такой случай, когда число текстов, релевантных для решения задачи об авторстве, больше 10. Естественным решением проблемы было бы объединение нескольких текстов в один, однако в случае спорного авторства это может оказаться невозможным. К тому же подобное решение нарушает выдвинутое Херданом положение о текстовом единстве как базе для выделения частей текста.

---

[27] G. Herdan. The Advanced Theory of Language as Choice and Chance, стр. 120.

Представляется, что проделанный нами обширный обзор статистико-лингвистической проблематики должен показать, что сейчас в центр лингвистической статистики входят две основные со всех точек зрения проблемы: это проблема случайности естественной языковой выборки и проблема стабильности лингвистических частот.

До работ Хердана эти проблемы почти не вставали, так как теоретико-статистический уровень работ по лингвистической статистике был слишком невысок. Сам Хердан принимал их как своего рода аксиомы, хотя развитие исследований по словарной статистике и заставляло его искать принципиально новые пути построения статистической модели словаря.

Мы исходим из того, что в настоящее время наиболее актуальной задачей лингвистической статистики является экспериментальная проверка на обширном материале этих двух наиболее существенных базисных статистических постулатов. Без подобной проверки движение лингвистической статистики вперед немыслимо.

Глава третья

ЭКСПЕРИМЕНТ ПО ПРОВЕРКЕ ОДНОРОДНОСТИ ПОЛЬСКИХ ТЕКСТОВ ОТНОСИТЕЛЬНО ЧАСТОТ ФОНОЛОГИЧЕСКОГО УРОВНЯ

§ 1. Методические вопросы. Некоторые проблемы польской фонологии

В качестве материала для статистического эксперимента нами были избраны тексты на польском языке. Польский язык был избран не случайно. Его положение в смысле статистической изученности являлось промежуточным. С одной стороны, существует всего лишь одно описание статистической структуры польской фонологии, выполненное в 1957 г. Марией Стеффен[1], так что имелась некоторая начальная система отсчета для возможной верификации наших результатов. С другой стороны, описание, данное в статье М. Стеффен, нельзя считать ни исчерпывающим, ни окончательным. К тому же в этой работе не ставились чисто статистические задачи, которые стоят в центре нашей работы. Поэтому казалось необходимым подвергнуть различные законченные тексты на польском языке статистическому обследованию с тем, чтобы установить, с одной стороны, определенные статистические параметры, могущие быть значимыми для языка в целом, а с другой стороны, проверить некоторые фундаментальные положения лингвистической статистики.

Далее, настоящая работа рассматривалась как определенное продолжение и развитие коллективной статьи группы авторов, предложенной в качестве доклада на V Международном съезде славистов в Софии [2]. В докладе ставилась задача построения типо-

---

[1] M. Steffen. Częstość występowania głosek polskich. «Biuletyn Polskiego Towarzystwa Językoznawczego», 1957, zesz. 16.

[2] М. И. Лекомцева, Д. М. Сегал, Т. М. Судник, С. М. Шур. Опыт построения фонологической типологии близкородственных языков. «Славянское языкознание». М., 1963.

логии близкородственных языков, причем таких, которые сохранили и типологическую близость. Был предложен определенный пересмотр традиционного понятия типа и введено понятие многоступенчатого типа. Сравниваемые славянские языки помещались в многопризнаковое пространство, и каждый язык получал определенную индексацию в терминах избранных типологических критериев. Собственно, задача состояла в том, чтобы получить осмысленное типологическое разделение весьма близких между собою языков. В частности, была разработана довольно подробная квалификация фонологических систем, являющаяся существенным развитием статистики фонемных инвентарей. Совершенно естественным кажется предпринять в продолжение подобной типологии сравнение языков на уровне статистических характеристик фонологической системы в текстах

Таким образом, перед нами стояла задача: выбрать определенное (достаточно большое) количество польских текстов и обследовать их статистически на уровне фонологии. С самого начала мы сознательно ограничили себя текстами, относящимися к художественной литературе. Почему было выбрано такое ограничение? К сожалению, оно во многом основывалось на соображениях технического удобства. Дело в том, что подавляющая часть элементарных подсчетов частот фонем и все статистические подсчеты (число последних доходило до десятков тысяч) были выполнены силами лишь автора настоящей работы, поэтому перед нами всегда стоял порог, перейти который было невозможно.

Мы вполне сознаем, что рассмотрение лишь одного из разнообразных возможных функциональных стилей языка суживает исследование и является определенным недостатком. Более того, мы предполагаем, что, возможно, разные функциональные стили являются взаимоисключающими генеральными совокупностями[3].

В силу ограниченности технических возможностей мы не могли обследовать в интересующем нас плане все функциональные стили польского литературного языка. Возник вопрос, какой из функциональных стилей следует обследовать статистически. Поскольку нас интересует статистика фонологических элементов, естественно было бы остановиться на так называемой устной речи: фонологическое исследование предполагает некоторое соотнесение со звучащей речью. Однако целый ряд трудностей заставил нас отказаться от этого намерения. С одной стороны, это были трудности чисто технического характера, связанные с тем, что нет изданных записей польской устной речи. С другой стороны, сама проблема установления того, что́ считать устной речью, чрезвычайно сложна. Очевидно, что это — не всякий текст, произносимый устно, не любая беседа (потому, что она может быть на весьма специализи-

[3] Ср.: Л. М. Гриднева. Розподіл голосних, приголосних і пропусків у сучасному українському мовленні. «Статистичні та структурні лінгвистичні моделі». Київ, 1966, стр. 53.

рованную тему), а некоторый особым образом определяемый текст, найти критерии которого чрезвычайно нелегко. Текст устной речи, который избирается объектом фоностатистики, должен быть максимально нейтральным, он должен «представлять» абстрактный польский язык. И здесь, за отсутствием подходящих текстов устной речи, мы решили избрать в качестве объекта анализа тексты польской художественной литературы. Нам представляется, что в некотором смысле стиль определенных видов художественной литературы (говоря в общем плане, то, что называют «беллетристикой») можно считать нейтральным, немаркированным. В подобном произведении автор стремится как можно точнее воспроизвести окружающую жизнь. Соответственно и в языковом плане мы будем иметь «слепок» с того языка, который общеупотребителен в жизни. Текст такого произведения нейтрален к прагматической функции (это — не заговор, заклинание или ораторский текст, где могут быть применимы специальные фонические приемы). Художественное произведение может включать в себя самую разнообразную лексику (в том числе и специализированную терминологию), в нем могут встречаться пассажи, максимально имитирующие устную, фамильярную речь, или речь официальную. При этом, поскольку художественное произведение является «естественной моделью» жизни, можно ожидать, что в его языке будут соблюдены истинные пропорции элементов различных функциональных стилей лучше, чем в случае искусственного подбора.

Короче говоря, художественное произведение, по крайней мере на фонологическом уровне, является неплохой «выборкой» из различных функциональных стилей языка. Наверное, художественный текст — наилучший «портрет» языка, на котором он написан.

Таким образом, польский язык, который обследуется в настоящей работе,— это язык художественной литературы. Интересующая нас задача формулируется в одном из своих аспектов как определение статистических свойств связных законченных текстов. Нас интересовало, насколько такие тексты однородны относительно частот фонологического уровня и насколько их можно считать случайными выборками из генеральной совокупности. Иными словами, нас интересовало, до какой степени реальные языковые тексты являются статистическими образованиями. Мы делаем упор на связности и законченности текста — качествах, необходимых для того, чтобы текст мог служить целям общения и передачи информации. С другой стороны, эти свойства явным образом противопоставлены нарочитой случайности, которая неизбежна, когда выборки образуются специально как набор независимых случайных величин. Было интересно выяснить, как связность и законченность текста влияют на свойства статистических параметров, находимых в предположении, что текст — это случайная выборка.

Естественно, что связные законченные тексты, избираемые в качестве базы подсчетов, должны быть не слишком велики. По-

мимо чисто технических условий обозримости подсчетов мы руководствовались следующими соображениями (которые, впрочем, достаточно спорны). Нам представлялось, что чем длиннее текст, тем меньше априори шансов на то, что он образует случайную выборку, и тем больше шансов на то, что он сам является генеральной совокупностью, четко отграниченной от других подобных совокупностей. Приведем следующее рассуждение. Допустим, нам дан минимально упорядоченный текст. (Мы имеем в виду упорядоченность, вносимую самой структурой текста, и в частности тем фактом, что текст является художественным произведением). Примером такого текста могут служить образованные случайным образом списки слов (некоторое подобие адресной или телефонной книги). Подобный текст можно без ущерба прервать в произвольном месте или, напротив, произвольно долго продолжать. Статистическая структура фонологического уровня будет минимально зависеть от длины текста.

Если же текст имеет определенную внутреннюю композицию, то его случайный характер уменьшается. Предполагается, что чем длиннее художественный текст, тем больше возможность внесения в него различного рода упорядоченности (разделение на содержательные части — главы, каждая из которых может оперировать своей лексикой; различная структура начала и конца текста,— ср. помещение в конец «Войны и мира» философских частей, статистическая структура которых даже на фонологическом уровне может отличаться от структуры других частей). Текст становится все менее однородным внутри себя самого. Неясно, насколько это рассуждение относится к фонологическому уровню (по отношению к лексическому и синтаксическому уровням оно справедливо, ср. данные Г. А. Лесскиса о различном синтаксическом строении определенных композиционных частей произведения). Однако, если композиция произведения такова, что различные его части строго отличаются друг от друга по теме, а каждая тема характеризуется определенной лексикой, то, очевидно, это может отразиться и на доле различных фонем в той или иной части текста.

Поэтому текст должен быть по крайней мере однотематичен и не содержать ясно выраженных композиционных членений. Подобным требованиям лучше всего отвечают рассказы и отдельные главы более крупных произведений. Брать в качестве отдельной выборки повесть или роман в свете вышеизложенного представляется неправильным.

С другой стороны, слишком малый объем текста также может влиять в сторону уменьшения возможности однородности. Эта опасность оказывается существенной именно для фонологического уровня. Если рассказ написан особой орнаментальной прозой, или если текст представляет собой стихотворение в прозе (примеры см. у Ю. Тувима: «Dziedzilija», «Zdrada», «Prowincja», «Przeklęty śpiew»), причем организация фонического уровня специально вхо-

дит в художественное задание, то повторяемость фонем в подобных текстах (вследствие повторов, аллитераций, ономатопэии) будет явно неслучайной. Таким образом, подобный малый текст никак нельзя считать случайной выборкой из общеязыковой совокупности.

Исследователь стоит между двумя опасностями: выбором текста, слишком сложно организованного и поэтому внутренне неоднородного, и выбором текста внутренне однородного, но резко отличающегося по своим статистическим характеристикам от любого другого текста.

Мы не претендуем на то, что наш выбор оказался оптимальным, во всяком случае, мы остановились на текстах, которые тяготеют к определенной нейтральности в смысле языковой организации.

1. Ярослав Ивашкевич. «Девушка и голуби» (Jarosław Iwaszkiewicz. «Dziewczyna i gołębie»). Глава 1. Повесть о современной Польше, написанная признанным мастером польской прозы. Язык повести прост. Композиционно глава является чисто повествовательной и сочетает авторскую речь (стилистически нейтральную, не обработанную под «сказ») с весьма простыми стилистически не маркированными диалогами, принадлежащими героям повести — молодым людям. Этот материал представляется нам наиболее нейтральным, ориентированным исключительно на повествование.

2. Леон Кручковский. «Первый день свободы». (Leon Kruczkowski. «Pierwszy dzień wolności»). 1-й акт. Взят 1-й акт пьесы о последних днях войны в Польше. Герои — солдаты и офицеры. Диалог может рассматриваться как хорошая модель непринужденной и даже фамильярной устной речи. Содержание диалога достаточно общее, неспециализированное. Текст был выбран как образец живого диалогического текста, описывающего внешнюю ситуацию, достаточно представительную как для польской литературы, так и для польской жизни последних тридцати пяти лет.

3. Ежи Шанявский (Jerzy Szaniawski). Пять рассказов из цикла «Profesor Tutka». Крайне простое и неспециализированное содержание. Рассказы напоминают притчи. Никакой ориентации на форму. Почти исключительно авторская речь, слегка стилизованная под речь «типичного» интеллигента. Сочетание живых речевых интонаций с тщательностью и обработанностью стиля.

4. Славомир Мрожек (Sławomir Mrożek). Пять рассказов из книги «Deszcz». Почти исключительно авторская речь. Сочетание общеупотребительных прозаизмов с «интеллигентными» и даже книжными оборотами. Стиль представляется нам достаточно типичным для письменной речи польских современных интеллигентов. Небольшая стилизация под «сказ». Богатая лексика.

Таким образом, мы сознательно выбрали польский язык в его литературном, «интеллигентном» варианте (хотя и с добавлением просторечий, что, впрочем, достаточно характерно для языка современных польских интеллигентов), а не в варианте «простонародном»

(ср. фельетоны и повести Веха-Вехецкого), ибо полагаем, что общее движение так называемого нейтрального стиля — именно в направлении сближения с речью интеллигентов. С одной стороны, мы стремились к тому, чтобы наши материалы были как можно более разнообразны, а с другой стороны, мы сознательно не включали те литературные явления, где присутствует нарочитая и сознательная ориентация на звуковую сторону (в русской литературе в качестве примеров можно было бы привести А. Белого или А. Веселого).

В какой степени наш выбор характеризует язык современной польской художественной литературы? По-видимому, в достаточно небольшой, поскольку эта литература весьма разнообразна и включает много направлений. Поскольку мы не ставили себе целью характеризовать художественные особенности современной польской литературы, а стремились найти репрезентанты некоторого нейтрального стиля, характеризующего польский язык «вообще», нам представляется, что этот выбор достаточно представителен.

* * *

После того, как были выбраны исходные тексты, возникла проблема перевода их в фонологическую транскрипцию. И здесь вопросы чисто теоретического, лингвистико-дескриптивного характера неизбежно переплетаются с проблемами техническими (удобство записи, сокращение предварительной технической работы и т. п.).

На первый взгляд может показаться, что необходимость фонологического транскрибирования возникает лишь потому, что мы имеем дело с письменными исходными текстами. Несомненно, однако, что это не так. Эта проблема должна возникнуть в любом случае, независимо от того, имеем мы дело с письменным текстом, или записываем устную речь на магнитофон. Различие лишь в том, что при переводе письменного текста в транскрипцию надлежит (как правило) расширить число употребляемых символов, а при переводе в транскрипцию устного текста обычно происходит сокращение числа символов за счет сведения разнообразных индивидуальных вариантов к единым типам.

В обоих случаях необходимо получить фонологически значимое разбиение текста на дискретные элементы, которые можно считать. Транскрипция, по нашему мнению, должна быть именно фонологической, а не фонетической, во-первых, потому, что мы ориентируемся на абстрактный фонологический уровень, а не на конкретный уровень фонетики и, следовательно, индивидуальной речи с присущим ей произношением. Транскрипция при фоностатистике должна представлять собою объективированное, обобщенное представление текста, а не запись реального произношения. Во-вторых, естественно стремиться к тому, чтобы наши результаты были сравнимы с результатами, полученными для других языков, имея в виду возможную типологию. Добиться этого можно лишь используя

не фонетическую транскрипцию, отражающую систему языка через сложную призму индивидуальных особенностей, но более обобщенную фонологическую транскрипцию.

При этом проблема сводится не только к технической стороне — переводу текста в дискретную последовательность фонологических символов, которые можно подсчитать. Гораздо более существенна другая сторона — определение того, какое из существующих фонологических описаний польского языка считать более адекватным с тем, чтобы избрать его в качестве базы для транскрипции.

И здесь мы оказываемся в центре весьма оживленной и еще далекой от окончания дискуссии о фонологическом составе польского языка. С самого начала следует заметить, что у обеих противоположных точек зрения, выявившихся в процессе дискуссии, есть, пожалуй, равное количество спорных и приемлемых моментов. Однако, в отличие от любого участника спора, нам пришлось не только выбирать ту или иную сторону в теоретическом плане (иногда даже пытаясь совместить и ту и другую точки зрения), но принять определенную систему фонологической записи в качестве основы для большой чисто практической работы. Поэтому мы не могли оставаться на позиции чисто теоретической оценки предложенных систем описания; необходимо было выбрать. Этот выбор, однако, не означает, что мы согласны со всеми теоретическими импликациями данной системы. Тем не менее он означает, что мы более склоняемся к одной точке зрения, чем к другой.

Изложим по возможности вкратце обе позиции. Расхождения между обеими точками зрения касаются сферы действия в польском языке дифференциальных признаков «носовость—ртовость» и «палатализованность—непалатализованность», а также состава вокализма, что непосредственно связано с обоими признаками. Оба признака оказываются связанными друг с другом так, что определенное суждение относительно одного из них ведет к определенному суждению относительно другого. Первая точка зрения состоит в том, что постулируется релевантность признака «палатализованность—непалатализованность» для большинства согласных фонем (есть некоторые вариации в отдельных решениях, сводящиеся к включению некоторых единичных фонем в сферу действия этого признака или исключению их из нее). Соответственно в составе вокализма выделяется лишь одна фонема [i], имеющая аллофоны [i] после мягких и [ɨ] (или в другой, более общепринятой в полонистике транскрипции [y]) — после твердых. Для этой же точки зрения характерно признание распространения признака «носовость—ртовость» на гласные при одновременном признании деназализации конечного носового *ẽ*.

Данная точка зрения на состав польских фонем была впервые высказана великим Бодуэном [4], и большинство польских фонетис-

[4] См.: «Materiały i prace Komisji Językowej», 1890—1891, v. I, стр. 134 и сл.

тов и фонологов ее придерживаются [5]. Однако уже в 1909 г. Казимеж Нитч выступил с весьма интересными возражениями против рассмотрения [i] и [y] как одной фонемы [6]. Далее эта точка зрения развивалась в направлении все большего сужения сферы признака «палатализованность—непалатализованность» в польском языке: в начале постулировалось отсутствие этого противопоставления в группе губных, затем в группе палатальных, наконец, в работе Виктора Яссема [7], одного из наиболее активных представителей этого течения, мы видим постулирование двух фонем [i] и [y] и решительное изгнание признака «палатализованность—непалатализованность» из системы польских согласных (равно как и признака «носовость—ртовость» из системы польских гласных).

Мы не будем анализировать отдельно каждую работу, в которой затрагивается проблема польских фонем. Попытаемся лишь суммарно представить аргументы обеих сторон.

Аргументы сторонников первой точки зрения относительно распространения в польском языке отношения «палатализованность—непалатализованность» и соотношения [i] — [y] можно свести к следующему. Польский язык рассматривается, во-первых, в свете своей истории и, во-вторых, в свете отношений, характерных для окружающих его родственных языков. Это — точка зрения, так сказать, извне. Действительно, для польского языка предшествующих периодов характерно распространение действия признака «палатализованность—непалатализованность». З. Штибер пишет в своей книге «Фонологическое развитие польского языка», резюмируя все развитие фонологической системы польского языка, следующее: «Сильное развитие палатализации согласных привело к значительному увеличению числа согласных фонем, а падение слабых еров привело к тому, что возникли разнообразные группы этих фонем» [8]. Имеется естественное стремление рассматривать современное состояние польской фонологической системы как результат тенденции предыдущего развития, а коль скоро вся эта тенденция говорит о нарастании на протяжении истории языка действия отношения «палатализованность— непалатализованность», то очевидно, что современные фонологические отношения должны отражать работу этой тенденции.

---

[5] См., в частности: S. S z o b e r. Gramatyka języka polskiego. Warszawa, 1959; Z. S t i e b e r. Rozwój fonologiczny języka polskiego. Warszawa, 1962; T. B e n n i. Fonetyka opisowa języka polskiego. Wrocław, 1959; E. S t a n k i e w i c z. The Phonemic Pattern of Polish Dialects. «For Roman Jakobson». The Hague, 1957, стр. 518 и сл., а также: С. К. Ш а у м я н. История системы дифференциальных элементов в польском языке. М., 1959.

[6] K. N i t s c h. Stosunek [i] do [y]: spółgłoski podniebienne i niepodniebienne. «Wybór pism polonistycznych». Wrocław, 1954.

[7] W. J a s s e m. The Distinctive Features and the Entropy of the Polish Phoneme System.— BPTJ, zesz. XXIV, 1966, стр. 87 и сл.

[8] Z. S t i e b e r. Указ. соч., стр. 87.

Другим вполне понятным аргументом является то, что польским палатализованным находится регулярное соответствие в русских мягких согласных (хотя возможны и отдельные расхождения, не нарушающие, впрочем, общей картины). Поскольку русский и польский языки в принципе демонстрируют действие одинаковой палатализационной тенденции после падения редуцированных, естественно рассматривать отношения между русскими мягкими и твердыми и польскими палатализованными и непалатализованными как отношения изоморфные. К тому же, как показывают работы Н. С. Трубецкого [9] и Р. О. Якобсона [10], отношение «палатализованность—непалатализованность» характерно именно для языков восточной Европы, и особенно для славянских, где оно является отличительным. Таким образом, по действию корреляции палатализации польский язык сближается не только с русским, но и с такими языками, как соседние — литовский, украинский и белорусский (при том, что в двух последних наблюдаются и диспалатализации), другие славянские — болгарский, македонский и лужицкие языки, сюда же входят территориально близкий румынский язык и мн. др. С точки зрения подобной языковой типологии отношения в польском ничем не отличаются от соответствующих отношений в других языках с категорией палатализации.

Аналогично решается и связанный с корреляцией палатализации вопрос об [i] — [y]. Еще К. Нитч в упоминавшейся статье об отношении [i] и [y] отмечает, что Бодуэн решает вопрос о польских [i] и [y] по аналогии с русскими *и* и *ы*. Действительно, с некоторой более общей точки зрения положение в обоих языках представляется аналогичным: в обоих случаях *ы* (*y*) не может выступать в начале слова, они сходны и дистрибутивно: перед *и* (*i*) выступают мягкие (палатализованные) согласные, перед *ы* (*y*) — твердые (непалатализованные). Если вспомнить, что в некоторых языках, например в эстонском, фонема, аналогичная *ы* (на письме обозначается как *õ*), свободно выступает в начале слова, а отношения «палатализованность — непалатализованность» отсутствуют, то положение в русском и польском покажется достаточно близким.

Вторая точка зрения стремится отказаться от исторических и типологических соображений и оперировать исключительно внутриязыковыми явлениями. Основные аргументы этого течения сводятся к отысканию таких фактов в польском языке, которые бы демонстрировали кардинальное отличие положения в польском от положения в других славянских языках (именно, в русском). Одновременно выдвигаются аргументы относительно того, что польские палатализованные (традиционно *ḿńṕb́v́f́śźćʒ́ḱǵ*, иногда *ẋ*) не представляют собою единого ряда, а являются классом гетеро-

---

[9] Н. С. Трубецкой. Основы фонологии. М., 1960, стр. 152—153.
[10] Р. О. Якобсон. К характеристике евразийского языкового союза. Париж, 1931.

генных элементов. Прежде всего указывают на то, что губные палатализованные в позиции перед любым гласным кроме *i* произносятся не как палатализованные + гласный, а как непалатализованные + *j* + гласный. Далее отмечается, что среднеязычные *śźćǵ* представляют собой не палатализованные согласные (соответственно утверждается, что у них нет непалатализованных соответствий), а собственно палатальные согласные. Что же касается *ḱ* и *ǵ*, то одни фонологи вовсе отрицают их существование (при этом положительно решается вопрос о существовании *ē* в том или ином виде на конце слова), другие же (В. Яссем) относят их к чисто палатальным согласным (соответственно для них вводятся обозначения *c* и *ɟ*). Остается *ń*. Здесь следует отметить, что вплоть до недавнего времени фонологи, придерживающиеся второй точки зрения, как бы забывали о том, что с изгнанием из системы бинарного противопоставления по «палатализованности — непалатализованности» все же остается одна фонема, для которой это противопоставление релевантно. Естественно, что иметь в системе признак, использующийся лишь для различения двух фонем, крайне неэкономно и неудобно. Поэтому в работах В. Яссема все доведено до конца и *ń*, обозначенное как ɲ, переведено в разряд палатальных.

Далее, требуется показать, что [i] и [y] не являются позиционными вариантами, а могут встречаться независимо. Здесь аргументы идут еще от К. Нитча, который отмечал следующие факты: [i] и [y] находятся не в дополнительном распределении, поскольку [i] встречается и после таких согласных, о которых утверждалось, что они могут предшествовать лишь [y]: *t*, *d*, *z*, *c*, *r*. Это происходит в следующих случаях: 1) в некоторых формах слов, которые, хотя и являются иностранными по происхождению, полностью ассимилированы польским языком, например kwestii ['kfesʦi], kurii ['kuri], pasii ['paśi], Irlandii [ir'landi] и т. п.; 2) в небольшом числе слов со старым *é* (т. е. полузакрытый передний гласный), как, например, zje [ʑi] (3 л. ед. ч. наст. вр. от глагола zjeść). Это, впрочем, характерно лишь для краковского диалекта, причем в его устаревшем варианте; 3) на границах морфем, как например в z innym ['ʑinnim], ziścić ['ʑiśćić]; 4) сюда же можно добавить такие, не отмеченные К. Нитчем, но по существу аналогичные случаи, как появление *i* после *z*, *s* и *r* в корневых морфемах некоторых иностранных слов, например Zanzibar, sinologia, cibazol, trik. Причем в последнем случае находится даже минимальное противопоставление trik [trik] 'трюк' — tryk [trɨk] 'баран'.

В. Яссем приводит дополнительные факты, позволяющие, по его мнению, утверждать фонологическое различие [i] и [y]: «1) Как *i* так и *ɨ* могут быть легко произнесены в изоляции любым поляком, владеющим литературным языком, в то время как в случае других гласных изолированно можно произнести лишь один вариант (так называемый главный член или основной вариант). Каждый учитель немецкого, французского или английского языка в Польше знает,

что различные виды *e* в этих языках (противопоставленные [ɛ]) очень трудно даются польским студентам, хотя в польском языке существует несколько разновидностей полузакрытых передних силлабических вокоидов — позиционных вариантов фонемы, главным членом которой является очень открытое [ɛ]. Напротив, противопоставление [i] — [ɨ] не вызывает никаких трудностей (например, в немецком или английском языке).

2) Большинство поляков называют букву *y* либо просто ɨ, либо ['ɨpsɨ͡ɔn] или ['ɨgrɛk]. Следовательно, для начальной позиции имеется противопоставление *i*—ɨ.

3) На протяжении почти четырехсот лет буква *i* последовательно обозначает [i], а *y* — [ɨ]. В современной лингвистике стало вполне общепринятым утверждение, что алфавитное письмо сознательно или бессознательно тяготеет к представлению фонем, а не аллофонов. Поскольку в весьма богатой фонемной системе польского языка два варианта одной фонемы никогда не различаются на письме, было бы странным, если бы это происходило в случае [i] и [ɨ]» [11].

Все указанные аргументы можно суммировать следующим образом: если стоять на точке зрения, что *i* и *y* — аллофоны, то следует выделить новый ряд палатализованных согласных *ʈdʐʂç* и т. п. в дополнение к традиционно выделяемым палатализованным. Однако поскольку подобное решение слишком неэкономно, целесообразнее ввести противопоставление фонем [*i*] и [*y*], что должным образом отразит наблюдаемые факты.

Каково наше отношение к изложенным точкам зрения по поводу состава польских фонем? Нам кажется, что обе отражают реальные факты; в каждой из предложенных схем акцентируется часть фактов. Несомненно, что К. Нитч и вслед за ним В. Яссем, П. Зволинский [12] и другие проделали очень важную и необходимую работу по установлению фонологических отношений в польском языке и выяснению того, как эти отношения отличаются от сходной на первый взгляд ситуации в русском и других славянских языках. Благодаря работам указанных лингвистов были вскрыты весьма тонкие фонологические отличия польского консонантизма от русского. Тем не менее мы не склонны считать проблему польских фонем решенной, как это делает В. Яссем. Это специфическое состояние нерешенности проблемы обуславливается, как нам кажется, реальным положением в языке, в котором происходят живые фонетические процессы, далекие, однако, от завершения.

---

[11] W. J a s s e m. A phonologic and acoustic classification of Polish vowels. «Zeitschrift für Phonetik und allgemeine Sprachwissenschaft», 1958, Bd 11, H. 4, стр. 303—304.

[12] P. Z w o l i ń s k i. Dokoła fonemów potencjalnych. «Lingua Posnaniensis», III, 1951; О н ж е. Stosunek fonemu [y] do [i] w historii języków słowiańskich. «Z polskich studiów sławistycznych». Warszawa, 1958.

Различные языки подвергаются фонологическому описанию на различных стадиях развития тех или иных фонетических процессов. Поэтому в одних языках (например, русском, где весьма велика унифицирующая роль литературного стандарта) оказывается легче установить стабильную систему фонем, охватывающую подавляющее большинство языковых фактов. В русском языке, по-видимому, в настоящее время не происходят в таком широком масштабе процессы, сходные с деназализацией конечного *ę* в польском и наблюдаемой в настоящее время тенденцией к его вторичной назализации под влиянием написания, что приводит к крайней вариативности, причем эта вариативность охватывает очень большой процент потенциального текста (русские примеры типа [что < што] характеризуют лишь периферию системы). В польском языке положение представляется не столь стабильным, как в русском, особенно в отношении категории палатализации. В русском языке палатализация выступает в наиболее четком, классическом виде. Палатализация охватывает практически всю систему согласных, и палатализованные могут находиться в любой позиции в слове. В польском языке произошла как бы сверхпалатализация. С одной стороны, это проявилось в переходе *r'* в *ž/š*, а с другой — в приобретении элементами *s'*, *z'* так называемого шипящего произношения [š'], [ž'] и в переходе *t'* и *d'* в аффрикаты [č'] и [dž']. Сюда же относится и «сверхотвердение» польского *ł* и переход его в [u̯]. Таким образом, польский язык, выражаясь фигурально, «проскочил» типологический этап, представленный русским типом палатализации. Всюду она не задержалась на «чистых» палатализованных, но пошла на шаг дальше: во всех палатализованных (а в *ł* — непалатализованном) появился дополнительный признак, осложнивший строение фонемы. Именно эта особенность является основным единым отличием польского консонантизма от русского. Насколько эта черта типична для ситуации в польском языке, видно из диалектного материала, например общепольск. *piwo* [pívo] — куявск. [pχívo] или [pš'ivo]. Правило добавления к палатализованному дополнительного фонетического признака здесь распространено и на губные. Палатализация как бы отделяется от палатализованного согласного и приобретает форму самостоятельного фонетического сегмента. Мы присутствуем, таким образом, при начале диспалатализации первичных палатализованных и образовании новых палатальных.

Встает, однако, вопрос, достаточно ли интенсивно проходил этот процесс, чтобы можно было утверждать, как это, по существу, делает В. Яссем, что диспалатализация в польском уже закончена и противопоставление палатализованных и непалатализованных уже более не релевантно. Приведем некоторые соображения по этому поводу.

1. Прежде всего отметим, что, по нашему мнению, центральный аргумент, выдвигаемый теми, кто придерживается точки зрения об отсутствии в польском языке противопоставления согласных по признаку «палатализованность — непалатализованность», допускает по крайней мере два толкования. Речь идет о так называемых губных палатализованных *ṕ b́ v́ f́ ḿ*. К. Нитч, а вслед за ним и некоторые другие польские фонетисты настаивают на том, что это не единые согласные фонемы, а сочетания с йотом. Действительно, аудиторный и акустический анализ показывает, что после *p b v f m*, когда они находятся перед любым гласным, кроме *i*, и являются «палатализованными», заслушивается явственный *i*-образный призвук, который виден на осциллограмме и связывает губной с последующим гласным. Таким образом, основной аргумент здесь — фонетический. Однако является ли этот призвук фонетически тождественным польскому йоту, который присутствует в слове самостоятельно (например, bajeczny, jajko и т. п.)? Те же экспериментальные материалы (например, Конечной и Скорупки[13]) показывают, что это не так. Фонетически *i*-образный призвук после палатализованных губных и *j* отличаются друг от друга. Если оставаться на тех же фонетических позициях, которые используются для обоснования «изгнания» губных палатализованных и введения вместо них сочетания «согласный +*j*», то следует использовать здесь не *j*, а ввести отдельную фонему [i̯]. Между прочим, говоря о появлении после польских губных палатализованных *i*-образного призвука, обычно забывают о том, что этот призвук зависит от последующего гласного, например, по нашим наблюдениям, он слабее перед [u] в таких словах, как miód, biuro, pióro и т. п., где он заслушивается крайне слабо и никак не может быть приравнен к *j*.

Основным доказательством отличия русских губных палатализованных от польских считается отсутствие *i*-образного перехода от согласного к последующему гласному в русском языке и невозможность в польском языке появления губных палатализованных в конечной позиции[14]. Относительно последнего трудно привести какие-либо контраргументы, но что касается первого доказательства, то в свете недавних исследований, Л. В. Бондарко и Л. Р. Зиндера[15], экспериментально доказавших наличие

[13] H. Konieczna. Przekroje rentgenograficzne głosek polskich. Wrocław, 1954; S. Skorupka. Studia nad budową akustyczną samogłosek polskich. Wrocław, 1955.

[14] См., в частности: W. Jassem. Рец. на книгу: M. Dłuska. Fonetyka Polska. Cz. I. Kraków, 1950. «Lingua posnaniensis», III. Poznań, 1951, стр. 323—339.

[15] Ср., в частности: Л. В. Бондарко, Л. Р. Зиндер. Дифференциальные признаки фонем и их физические характеристики. «XIII Международный психологический конгресс. Москва, 1966. Симпозиум 23. Модели восприятия речи». Л., 1966, стр. 36: «Наиболее общим признаком мягкости согласных является специфический *i*-образный переход следующего гласного: F_2 изменяется от 2000—3000 гц до частоты, характеризующей F_2 гласного».

совершенно аналогичного явления в русском языке, появление [i̯] после губных палатализованных перед гласными не может считаться аргументом против существования в польском губных палатализованных.

Отметим здесь, что еще Н. С. Трубецкой понимал палатализацию достаточно широко и уверенно включал явления, подобные появлению после губных палатализованных в польском языке, в категорию палатализации: «Отдельные языки, в которых эта корреляция встречается, весьма отличаются друг от друга также и в отношении фонетической реализации палатализованных согласных. Однако принцип всюду остается одним и тем же: «палатализованный» согласный имеет *i*- (или *j*-)образную окраску, которая сочетается с другими признаками этого согласного, тогда как соответствующий ему «непалатализованный» согласный лишен этой окраски; *i*-образная окраска палатализованного согласного возникает благодаря подъему средней части языка к твердому нёбу; чтобы особенно резко оттенить это противоположение, задняя часть языка при образовании непалатализованных согласных часто приподнимается к мягкому нёбу»[16].

Таким образом, в свете того, что *i*-образный призвук, по-видимому, неотделим от палатализации, следует констатировать, что польские губные *p' b' v' f' m'* вполне можно трактовать и с фонетической точки зрения как палатализованные.

2. Существует важный дистрибутивный аргумент в пользу введения сочетаний «согласный $+j$» вместо палатализованных: утверждается, что в польском языке не существует сочетаний *p b v f m* $+ j$, противопоставленных *p' b' v' f' m'*. Согласно этой точке зрения нет противопоставления между парами hrabia и Arabja, Ziemia и Warmja.

Помимо аудиторных данных сюда привлекаются и данные по польской рифме[17]. Показывается, что, начиная с XVII в., в польской поэзии зарегистрированы случаи, когда в одной из двух рифмующихся строк последний слог строится как сочетание «согласный $+ j +$ гласный», а в другой — «согласный $+$ гласный»; например, Потоцкий (XVII в.): ceremonii — dłoni; pani — w plebanii. Особенно много подобных примеров у Словацкого: conwulsyi — ekspulsyi — czulsi; Konfederacyi — kaci — Wypłaci. Вообще подобный способ рифмовки характерен для поэзии XIX в. (есть многочисленные примеры у Гарчинского, Тетмайера, Каспровича).

К. Нитч делает на основании этих данных вывод о том, что элиминация *j* в сочетаниях «согласный $+ j +$ гласный» в окон-

---

[16] Н. С. Трубецкой. Основы фонологии, стр. 152—153. Ср. веляризацию *ł* в польском языке, приведшую к его замене на [u̯].

[17] K. Nitsch. Z historii rymów polskich. «Wybór pism polonistycznych». Wrocław, 1954, стр. 34—37.

чаниях заимствованных слов является закономерной тенденцией в развитии польского консонантизма. Согласно мнению К. Нитча, следует ожидать, что в дальнейшем эта тенденция будет развиваться.

Мы, со своей стороны, решили проверить, имеет ли тенденция к идентификации последовательностей C + *j* + V и C + V развитие в современной польской поэзии. С этой целью были просмотрены стихотворения Ю. Тувима (по собранию сочинений, вышедшему в Варшаве в 50-х годах), К. Галчинского (аналогично), Л. Слонимского (по однотомнику) и С. Гроховяка (также по однотомнику). (Стихи современных более молодых поэтов не годятся, так как написаны в основном без рифмы). Следует отметить, что по мере того, как поэзия становится более удаленной от моделей, представленных такими поэтами, как Словацкий, Каспрович и Тетмайер, уменьшается встречаемость рифм типа C + *j* + V = CV. Наибольшее количество их зафиксировано у Слонимского, ближе всего стоящего к традиции XIX в. (например, piersi — Persji, Arabii — krwawi, ziemi — alchemii, pasji — nasyp), у Гроховяка нет ни одного подобного случая, у Тувима встречаются даже специальные подчеркивания *j* в данной позиции: Konwulsji — puls jej. В остальных случаях (и у Тувима, и у Галчинского) имеется соответствие C*j*V = C*j*V (например, Konsultacje — rezygnacje), не являющееся диагностическим.

По-видимому, в общем плане можно говорить о том, что тенденция к элиминации *j* в окончаниях, отчетливо выступавшая в XIXв., в XX в. затормозилась. Во всяком случае, нельзя говорить о том, что окончательно победила тенденция к устранению *j*. По нашим собственным наблюдениям, в речи образованных поляков скорее проявляется стремление сохранить *j* в подобных случаях. Происходит это, по-видимому, под влиянием правописания в связи с широким распространением грамотности. К тому же полному устранению *j* в словах типа chemia, Arabja мешает то, что аналогичные окончания «*j* + гласный» зафиксированы и в случаях, где устранение *j* должно привести к образованию «нового палатализованного» (например, Rosja, Arkadja и т. п.), чему язык все-таки противится, несмотря на спорадические факультативные случаи типа родительного падежа единственного числа kurii, Anglii и т. п. Полному устранению *j*, наконец, мешает и то, что есть случаи типа Danja, где выпадение *j* должно было бы привести к произношению ['Dańa], что воспринимается как явно недопустимое и нелитературное.

С другой стороны, сама аргументация отсутствия губных палатализованных снятием противопоставления между губными палатализованными в случаях типа biały, hrabia и случаях типа Bianka, Arabja в результате устранения *j* в последних двух словах не подтверждается, а скорее опровергает исходную гипоте-

зу. Предположим, что в случаях Bianka, Arabja *j* действительно устранен. Тогда получаем две ситуации: в словах biały, hrabia сохраняются губные палатализованные, а в словах Bianka, Arabja они появляются в результате устранения *j*. Таким образом, здесь мы действительно находим нейтрализацию противопоставления «палатализованный — непалатализованный», но такую, при которой победителем выходит именно палатализованный член корреляции.

По нашему мнению, именно в этом и заключается тенденция развития консонантных сочетаний с *j* в заимствованных словах. Как пишет З. Фолеевский: «Тенденция отождествлять в произношении группы *bj*, *pj*, *vj*, *fj*, *mj* (а также *nj*) и *b'*, *p'*, *v'*, *f'*, *m'*, (и *n'*) не является аргументом в пользу того, что палатализованные согласные *b'*, *p'*, *v'*, *f'*, и *m'* следует считать двухфонемными сочетаниями [b] + [j], [p] + [j], [v] + [j], [f] + [j], [m] + + [j]. В таком случае следовало бы палатальное [ń] (или [ɲ]) считать сочетанием [n] + [j]. Напротив, сочетания [согласных с *j*— *Д. С.*] (первоначально обнаруживаемые в срединной позиции в заимствованиях и не существовавшие в собственно польских словах, за исключением позиции стыка морфем) часто отождествляются с отдельными согласными»[18].

Еще раз повторим, что, по нашему мнению, тенденция к отождествлению групп «согласный + *j*» с одним палатализованным согласным полностью никогда не осуществляется, так как это привело бы к возникновению в польском языке длинного ряда вторичных палатализованных *t'*, *d'*, *c'*, *z'*, *r'*, *s'* и т. п. Вторая гипотетическая ситуация, возникающая при устранении *j* в сочетаниях «согласный + *j*» в заимствованных словах, представляется совершенно неправдоподобной — в словах biały, piasek *j* появляется, а в словах Arabja, chemia — исчезает. Подобное явление было бы совершенно необъяснимо фонетически, а, кроме того, не свидетельствовало бы об отсутствии противопоставления по «палатализованности—непалатализованности» у губных, так как ликвидация губных палатализованных в первом случае компенсировалась бы их возникновением во втором.

Таким образом, мы видим, что «дейотизация» сочетаний «согласный + *j*» в заимствованных словах не может служить доказательством отсутствия противопоставления по «палатализованности—непалатализованности» у губных.

Остается рассмотреть действительно имеющую место ситуацию фонетического сближения случаев biały и Bianka, состоящего в том, что в обоих случаях после *b* присутствует сходный *i*-образный призвук. Мы, однако, предпочитаем считать, что заимствованные слова ассимилируются под исконные, а не наобо-

---

[18] Z. Folejewski. The problem of Polish phonemes. «Scando-Slavica», t. II.

рот, и полагаем, что не biały стремится к ['bjały] и, следовательно, к ['Bjanka], но Bjanka стремится к ['Bi̯anka] и, следовательно, к ['b'ały]. В современной речи, как мы уже упоминали, существует тенденция, препятствующая полному слиянию обоих типов.

Итак, если стоять на фонетической точке зрения, то следует признать, что в реальной речи происходит отождествление сочетаний «согласный + *j*» с палатализованными, а не наоборот. Если же избрать фонологическую точку зрения, то следует указать, что в парадигматическом плане это отождествление еще не произошло и, следовательно, имеются позиции противопоставления *b' p' m' f' v'* сочетаниям *bj pj mj fj vj*.

3. Другим аргументом, выдвигаемым в пользу разделения фонем [i] и [y], являются случаи возникновения так называемых вторичных палатализованных *t' d' r' z' s' c'* и т. п. Здесь происходит определенное смещение акцентов. Все описываемые случаи зарегистрированы в заимствованных словах и могут рассматриваться как сигналы чуждой фонологической системы [19]. Против этого можно выставить аргумент, что подобным сигналом следует считать сегменты, идентифицируемые различительными признаками, отсутствующими в первоначальном наборе для данного языка (ср. произношение фамилии известного польского лингвиста A. Brückner'a: *brykner* или *br'ikner*, но не *brükner*). Во всяком случае, приводимые примеры относятся исключительно к периферии фонологической системы и возникающие *t' d' s' z'* и т. п. не могут быть приравнены к фонемам, образующим четкие парадигматические ряды. Так называемые вторичные палатализованные возникают либо в сочетании с *j* (Francja, Rosja, Kurja и т. п.), либо в изолированных, единичных случаях, не образующих устойчивых рядов, а существующих лишь в виде отдельных примеров. Относительно случая с *j* мы высказали свое мнение выше — фонологически их следует считать сочетаниями непалатализованных согласных с последующим *j*, противопоставленных палатализованным согласным (по типу ['xrab'a] — [a'rabja]). Что же касается случаев типа trick [trik], cibazol, sinus, то представляется, что не имеет смысла вводить отдельный ряд вторичных палатализованных для учета этих случаев, так как их число в лучшем случае достигает 10—15 (кстати, все авторы, пользующиеся данным аргументом в защиту расщепления [i] и [y], обращаются к одним и тем же немногим примерам). Если ввести разделение [i] и [y] для учета этих случаев, то в русском языке, следуя аналогичной логике, нужно было бы ввести [h] для немецких слов типа *Гейне* ['hainə] или [ø] для слова *шедевр* [ʃədøvr] в некоторых произношениях.

---

[19] Ср.: Z. Stieber. O zaburzeniach równowagi fonologicznej.— BPTJ, zesz. IX, 1949, стр. 79—81.

Так же следует трактовать приводимые В. Яссемом примеры появления *y* в начале слова в словах ['ɨpsɨlõn] или ['ɨgrek] — всего два слова во всем языке. Подобным образом число фонем можно увеличивать почти ad infinitum.

4. Пожалуй, единственный аргумент в пользу разделения [i] и [y], который не теряет своей значимости,— это соображение относительно появления вторичных палатализованных в проклитиках перед словами, начинающимися с *i*. Дело в том, что при транскрипции текста неизбежно встает вопрос о передаче явлений сандхи, возникающих внутри слова и на его границах. Неизбежно приходится выбирать основную синтагматическую единицу, внутри которой допускаются явления сандхи, поскольку рассматривать подобные явления в достаточно больших отрезках текста (например, предложениях) крайне затруднительно, во-первых, потому, что это связано с огромной работой по перекодированию текста и, во-вторых, поскольку за пределами слова эти явления теряют свой обязательный характер и переходят на уровень индивидуальной речи. Мы могли учитывать явления сандхи только в тех случаях, где они совершенно обязательны, т. е. являются фактами языка. Обязательный характер явления сандхи носят лишь внутри слова, а также вне его в случаях примыкания проклитик и энклитик.

Так возникает в качестве основной синтагматической единицы, внутри которой происходят фонетические процессы, ф о н е т и ч е с к о е с л о в о, представляющее собою полноударное слово с примыкающими к нему проклитиками или энклитиками. К проклитикам относятся предлоги, союзы и т. п., к энклитикам — частица się и краткие формы личных местоимений mu, go, mi и т. п.[20] Внутри подобного фонетического слова учитывались при транскрипции регрессивное оглушение и озвончение согласных (например, nad stołem [natstołem]; w pamięci [fpaṁeńći], k domu [gdomu]; прогрессивное оглушение *v* после глухих (kwiat [kf́at]); оглушение согласных на конце слова (rycerz [riceš]), а также смягчение проклитик перед *i* в начале слова (w Instytucie = [vinstituće]).

Естественно, что проблема *d* и *z* перед *i* (другие «непалатализуемые» фонемы в проклитиках в этом положении не встречаются) представляет реальную трудность. Мы решили в таких случаях транскрибировать *d*/*z* + *ji*-, например *z* innym [zjinnim]. Подобное решение было выбрано, во-первых, потому, что, по нашим наблюдениям, вставление *j* в таких случаях характерно для произношения самых различных носителей польского языка и является весьма распространенным; во-вторых, это позволяет избежать введения вторичных палатализованных.

---

[20] Полный список см. в книге: S. S z o b e r. Gramatyka języka polskiego. Warszawa, 1959, стр. 22—24.

Мы вполне сознаем условный характер введенной нами единицы — фонетического слова. По-видимому, в реальной речи и обязательные сандхи будут носить иной характер, чем в подобной записи, в частности, оглушений согласных на конце слова будет явно меньше, поскольку в реальной речи в единое фонетическое целое могут стягиваться не только полноударные слова вместе с энклитиками и / или проклитиками, но и несколько полноударных слов.

Однако мы стремимся избрать как можно более объективный способ репрезентации текста, а этого можно было достичь, учитывая лишь процессы, которые будут происходить всегда.

Таким образом, мы делаем вывод, что и внутри фонетического слова на границе проклитик вторичные палатализованные *t' d' s' z' c'* и т. п. являются свободными вариантами соответствующих непалатализованных фонем и возникают спорадически на периферии системы. Отпадает еще один аргумент в пользу разделения фонем [i] и [y] и против признания в польском языке палатализованных *ṕ b́ ŕ v́ ḿ ń ś ź ć ǵ ḱ ǵ*. По нашему мнению, даже признание возникновения в указанных позициях вторичных палатализованных не является абсолютным аргументом в пользу разделения [i] и [y]. Авторы, стоящие на точке зрения, отрицающей релевантность категории палатализации для польского языка, не замечают, что устраненная из одной части системы эта категория проникает в другую его часть. Они объясняют различие между словами kurii и kury как [kuri] и [kury], но сами говорят о том, что в слове kurii возникает палатализованный *r'*. Поскольку в принципе возможно различение в языке узкого и широкого [i] без смягчения согласного перед узким [i] (например, в английском и немецком языках), неясно, почему в данном случае палатализацию надо относить к гласному.

Исходя из приведенного разбора аргументов против введения в польском языке категории палатализации при описании фонологической системы, мы приходим к выводу, что эти аргументы не имеют решающего характера и что категория палатализации должна быть сохранена, хотя несомненно, что ее характер отличается от русской палатализации.

Теперь рассмотрим положение с категорией «носовость—ртовость» в системе польских гласных. Параллельно с устранением категории палатализации из польского консонантизма в последних работах В. Яссема[21], а также Л. Беджицкого[22] утверждается, что «процесс устранения носовых гласных фонем закончился. В современном литературном польском языке носовые гласные

[21] W. J a s s e m. The distinctive features and the entropy of the Polish phoneme system.— BPTJ, zesz. XXIV, 1966, стр. 87—108.

[22] L. B i e d r z y c k i. Fonologiczna interpretacja polskich głosek nosowych.— BPTJ, zesz. XXII, 1963, стр. 25—45.

функционируют не как самостоятельные фонемы, а как позиционные варианты гласных фонем, главные варианты которых ртовые»[23].

С носовыми гласными дело обстоит иначе, чем с категорией палатализации у согласных. В целом дистрибуция палатализованных согласных свободнее, чем дистрибуция носовых гласных, хотя для определения категорий палатализованных (губные) конечная позиция и является запрещенной. Традиционно отмечается, что носовые гласные встречаются лишь в строго ограниченных контекстах (õ на конце слова и в середине перед шипящими и свистящими *s ś š z ź ž* и перед *x*, а *ẽ* лишь перед шипящими, свистящими и *x*). В других контекстах носовые гласные либо вовсе утрачивают носовость (конечное *ẽ*), либо передают ее специальному согласному, появляющемуся между деназализованным гласным и последующим согласным, например węgiel, dąb = [veŋǵel], [domp].

С другой стороны, встречаются сочетания гласных с носовыми согласными перед шипящими и свистящими в заимствованных словах, например awans, inspektor, kunszt и т. п., в которых произносятся новые носовые гласные *ã ĩ ũ* и т. п. Отличие ситуации носовых гласных от ситуации палатализованных согласных представляется нам существенным. Оно состоит в том, что так называемые губные палатализованные, по нашему мнению, еще нельзя считать окончательно диспалатализовавшимися и расщепившимися на согласный и *j*, в то время как носовые гласные перед смычными согласными давно уже расщепились на ртовый гласный и носовой согласный.

Обе категории отражают действие одной тенденции, но в случае носовых гласных эта тенденция зашла дальше. Это дает основание Л. Беджицкому и В. Яссему утверждать, что выделение признака «носовость» в отдельную фонему произошло не только перед смычными, но и во всех позициях, т. е. и в конечной (для *õ*); таким образом, слово chodzą в транскрипции В. Яссема будет выглядеть как [xodzoŋ]. В. Яссем выделяет следующие носовые фонемы (согласные!): [m], которая помимо случаев, обозначенных буквой *m*, представляет также случаи типа gęba [gemba], rąbie [roḿbe] — т. е. «носовые гласные» перед губными смычными; [n], которая помимо звуков, обозначенных буквой *n*, представляет также «носовые гласные» перед *t d cz čž*, например tędy [tendɨ], kąt [kont], pączek [ponček], ręce [rence], prędzej [prenzej]; [ɲ], которая представляет звуки, обозначаемые буквой *ń*, а также *n* перед *i*, например koń [koɲ], goni [goɲi], koński [koɲsci] (транскрипция В. Яссема), и «носовые гласные» — перед *ś ź ć ǯ*, например bądź [boɲć], pięć [pjeɲć], pięść [pjeɲść], więź [vjeɲś], więzią [vjeɲźoŋ], и т. п., а также [ŋ], обозначающая носовые гласные

[23] Там же, стр. 40.

во всех прочих случаях, т. е. перед *s z š ž*, например kęs [keŋs], mięso [mjeŋso], rzęsy [žeŋsɨ], wąs [voŋs], mąż [moŋš] и т. п., перед *k g*, например mąka [moŋka], tęgi [teŋgi], bank [baŋk] и т. п., и на конце слова: są [soŋ], idą [idoŋ], в некоторых идиолектах idę [ideŋ], się [śeŋ] и т. п.

Подобное решение проблемы носовых гласных в польском языке несомненно является весьма последовательным и радикальным. Положительной стороной такого фонологического описания подсистемы носовых является устранение асимметричности в вокализме. При введении в польский вокализм носовых фонем исследователь обычно стоит перед следующей дилеммой: выделять ли два носовых $\tilde{o}$ и $\tilde{e}$ или лишь один $\tilde{o}$. С одной стороны, статус $\tilde{o}$ и $\tilde{e}$ несомненно неодинаков: дистрибуция $\tilde{e}$ гораздо более ограничена, чем дистрибуция $\tilde{o}$, и в чистом виде $\tilde{e}$ встречается лишь в единичных словах (kęs, rzęsy и т. п.), поэтому некоторые фонологи (например, З. Штибер) считают, что $\tilde{e}$ является своего рода «потенциальной» (точнее, маргинальной) фонемой и его не следует включать в основной инвентарь фонем, как не включаются в него $\tilde{a}$ $\tilde{u}$ $\tilde{i}$ и т. д. Однако исключение $\tilde{e}$ из числа фонем приводит к существенной деформации системы, поскольку типологически система с одним носовым гласным должна отличаться от того, что имеет место в польском языке [24].

Поэтому введение вместо «носовых гласных» во всех позициях сочетания ртового гласного с носовым согласным снимает проблему потенциальной фонемы $\tilde{e}$. Признак носовости выделяется в отдельную согласную фонему. Система носовых согласных m nŋ̣ обладает внутренней стройностью и представляется в общем убедительной. В теоретическом плане подобное описание подсистемы носовых более приемлемо, чем описание подсистемы палатализованных у В. Яссема, поскольку, как уже указывалось, носовые гласные представляют собою гораздо более эволюционировавший фрагмент системы, чем палатализованные согласные.

Однако мы решили остановиться на более традиционном описании системы носовых, которое выделяет согласные *m ḿ n ń* и гласные $\tilde{o}$ и $\tilde{e}$. Это объясняется прежде всего тем, что, поскольку мы принимаем существование в польском языке категории «палатализованных», фонема, которая у В. Яссема обозначается как ŋ̣, у нас представлена как палатализованное *ń*. На первый взгляд

---

[24] Ср.: Н. С. Трубецкой. Основы фонологии, стр. 139: «Окраска такого единственного назализованного гласного определяется консонантным окружением, а его степень раствора как признак вообще не существует. Иными словами, этот «неопределенный» назализованный гласный является не чем иным, как слоговым носовым, артикуляция которого ассимилируется (уподобляется) артикуляции последующего согласного». Это описание противоречит статусу польского носового $\tilde{o}$, если лишь его выделять как единственный носовой гласный.

безразлично, как обозначать одно и то же явление. Однако это не так. У В. Яссема ɲ в качестве бинарного минимального противопоставления имеет фонему ŋ, которой он противопоставлен как «острый» (acute или, по другой терминологии, «непериферийный») «тупому» (grave или, по другой терминологии, «периферийный»). В системе противопоставлений, использующейся нами, *ń* отождествляется как имеющий «плюс» по признаку «палатализованность», а фонемой, получающей по этому признаку «минус», является *n*. Таким образом, если бы в нашей системе была сохранена фонема ŋ, она была бы исключена из непосредственного бинарного соотношения с фонемой *ń*.

В. Яссем вводит ɲ, стремясь исключить категорию «палатализованности» из системы согласных. Эта фонема ставится В. Яссемом в ряд палатальных (наряду с *ç ʒ̨ t̨ ç̣ d̨ ʒ̣* — транскрипция Яссема, у нас — *ć ź ć̣ ʒ́*), однако в матрице идентификации польских фонем, приводимой в статье Яссема [25], никакой различительный признак не объединяет эти фонемы (*ń ś ź ć ʒ́*), характеризующиеся несомненным фонологическим и фонетическим единством, в одну группу. Таким образом, в системе Яссема фонема ɲ необоснованно выделяется из всех «бывших палатализованных»: то, что раньше считалось *ń*, у Яссема ни разлагается на *n* и *j* (как в случае губных), ни объединяется с палатальными. Таким образом, решение, принятое В. Яссемом, вовсе не свободно от неясностей.

Поэтому мы сочли более целесообразным оставить *ń* в ряду палатализованных, куда мы включаем и *ś ź ć ʒ́*.

Поскольку в нашей системе вместо ɲ принято традиционное *ń*, возникает вопрос, как поступить с ŋ. В системе В. Яссема оно играет большую роль, так как обозначает носовость там, где обычно выступают носовые гласные.

Так же, как и в случае губных палатализованных, здесь, по нашему мнению, происходит переоценка фонетических фактов: для губных *i*-образный призвук, присущий любой палатализации, интерпретируется как отдельная фонема *j*, а для носовых в конечной позиции и перед шипящими и свистящими глайдообразный призвук *w̃* (*j̃* перед палатализованными *ś ź*), например *są*[*sow̃*], mięsisty[mẽj̃śisti], который можно трактовать как сопутствующий, трактуется как отдельная фонема ŋ (или ɲ).

Против трактовки этого призвука как отдельной консонантной фонемы можно привести следующие аргументы. Во-первых, чисто фонетический аргумент, что в конечной позиции *õ* заслушивается все-таки как *õ*, а не как *oŋ*. Если бы носовой элемент представлял собою отдельную согласную фонему, то в сочетании конечного *õ* с последующим гласным ясно выступила бы фонема ŋ. Этого же не происходит, например, są oczy [sõ ̯oči], а не [sõŋ ̯oči].

---

[25] W. J a s s e m. The distinctive features and the entropy of the Polish phoneme system, стр. 102.

Еще один аргумент типологического плана. В языках, имеющих фонему ŋ (с ограничением дистрибуции: начальная позиция запрещена), как правило, обнаруживается придыхательная фонема *h*, с которой ŋ находится в дополнительном распределении,— в польском языке такой фонемы нет. Далее, если бы ɲ было полным коррелятом ŋ, можно было бы ожидать, что они будут встречаться в одних и тех же позициях, чего не происходит: ɲ встречается в начальной позиции.

Наконец, отметим, что если выделение *j* после губных палатализованных может указывать на направление тенденции развития (ср. диалектные факты), то выделение *ŋ* в конечной позиции на подобную тенденцию не указывает (ср. полную деназализацию конечного *õ* в некоторых польских диалектах (средняя и северная Малопольша, Прикарпатье) [26].

Исходя из этих соображений, мы приняли традиционную схему ринезма польского литературного языка: *mṁnńõẽ*. Носовые гласные выделяются — *õ* в конечной позиции и перед свистящими, шипящими и *x*, *ẽ* только перед свистящими, шипящими и *x*. В остальных позициях они расщепляются на *o* или *e* и носовой согласный.

В конечном итоге фонологическая система польского языка принимает тот вид, который она имеет в коллективном докладе группы авторов на съезде славистов в Софии [27]. Для идентификации фонем используются следующие различительные признаки, введенные и усовершенствованные Р. Якобсоном, Г. Фантом и М. Халле: 1) гласность—негласность; 2) согласность—несогласность; 3) компактность—некомпактность; 4) диффузность—недиффузность; 5) периферийность—непериферийность; 6) непрерывность—прерывность; 7) назальность—неназальность; 8) яркость—тусклость; 9) звонкость—глухость; 10) палатальность—непалатальность. О значении отдельных признаков, а также символов идентификации «+», «—» и «0»— см. в упомянутом коллективном докладе (см. табл. 1).

Мы обсудили основные спорные моменты в фонологии польского языка. Следует подчеркнуть, что работы К. Нитча, В.Яссема и других имеют большую ценность, поскольку показывают специфику польской фонологии и указывают на некоторые тенденции ее развития. Тот факт, что мы избрали другую систему описания, говорит о том, что, по нашему мнению, еще рано утверждать, что изменения, постулируемые в работах этих ученых, уже произошли. К тому же выбор традиционной системы описания, в которой

---

[26] С. М. Толстая. К типологической интерпретации польского ринезма. «Лингвистические исследования по общей и славянской типологии». М., 1966, стр. 126.— В этой статье имеются ценные соображения о структуре носовых в польских диалектах и литературном языке.

[27] М. И. Лекомцева, Д. М. Сегал, Т. М. Судник, С. М. Шур. Указ. соч., стр. 454.

Таблица 1

Матрица идентификации фонем польского языка

Признак	*r*	*n*	*ń*	*ɥ*	*l*	*m*	*ḿ*	*j*	*t*	*d*	*s*	*ʒ*	*ʂ*	*ʐ*	*p*
Гласность	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Согласность	—	—	—	—	—	—	—	—	+	+	+	+	+	+	+
Компактность	—	—	—	—	—	—	—	+	—	—	—	—	—	—	—
Диффузность	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Периферийность	—	—	—	—	—	+	+	0	—	—	—	—	—	—	—
Непрерывность	—	—	—	+	+	0	0	0	—	—	—	—	+	+	—
Назальность	—	+	+	0	0	0	0	0	0	0	0	0	0	0	0
Яркость	0	0	0	0	0	0	0	0	—	—	+	+	0	0	0
Звонкость	0	0	0	0	0	0	0	0	—	+	—	+	—	+	—
Палатальность	0	—	+	—	+	—	+	0	0	0	0	0	0	0	—

Признак	*ṕ*	*b*	*b́*	*f*	*f́*	*v*	*v́*	*š*	*ś*	*ž*	*ź*	*č*	*ć*
Гласность	—	—	—	—	—	—	—	—	—	—	—	—	—
Согласность	+	+	+	+	+	+	+	+	+	+	+	+	+
Компактность	—	—	—	—	—	—	—	+	+	+	+	+	+
Диффузность	0	0	0	0	0	0	0	0	0	0	0	0	0
Периферийность	+	+	+	+	+	+	+	—	—	—	—	—	—
Непрерывность	—	—	—	+	+	+	+	0	0	0	0	0	0
Назальность	0	0	0	0	0	0	0	0	0	0	0	0	0
Яркость	0	0	0	0	0	0	0	—	—	—	—	+	+
Звонкость	—	+	+	—	—	+	+	—	—	+	+	—	—
Палатальность	+	—	+	—	+	—	+	—	+	—	+	—	+

Признак	*ǯ*	*ʑ*	*k*	*ḱ*	*g*	*ǵ*	*x*	*e*	*ẽ*	*o*	*õ*	*u*	*i*	*a*
Гласность	—	—	—	—	—	—	—	+	+	+	+	+	+	+
Согласность	+	+	+	+	+	+	+	0	0	0	0	0	0	0
Компактность	+	+	+	+	+	+	+	—	—	—	—	—	—	+
Диффузность	0	0	0	0	0	0	0	—	—	—	—	+	+	0
Периферийность	—	—	+	+	+	+	+	—	—	+	+	+	—	0
Непрерывность	0	0	—	—	—	—	+	0	0	0	0	0	0	0
Назальность	0	0	0	0	0	0	0	—	+	—	+	0	0	0
Яркость	+	+	0	0	0	0	0	0	0	0	0	0	0	0
Звонкость	+	+	—	—	+	+	0	0	0	0	0	0	0	0
Палатальность	—	+	—	+	—	+	0	0	0	0	0	0	0	0

используются различительные признаки Р. Якобсона и других, позволяет сравнивать это описание с фонологическими описаниями других языков, использующими те же различительные признаки.

В. Яссем, предложивший для польского языка систему фонем, в которой отсутствует противопоставление по палатализованности—непалатализованности, пошел в этом направлении еще дальше и выдвинул собственную систему различительных признаков, существенно отличающуюся от схемы Р. Якобсона [28]. Для идентификации 37 фонем В. Яссемом используются следующие 9 признаков: 1) консонантность—вокальность; 2) супраглоттальность—инфраглоттальность (supraglottal—infraglottal); 3) назальность—неназальность; 4) гладкость—прерывность (smooth—abrupt); 5) компактность—диффузность); 6) острый—тупой (acute—grave); 7) низкая тональность—высокая тональность (low-tone—high-tone); 8) краткий—долгий и 9) звонкий—глухой (см. табл. 2).

В. Яссем строит систему признаков исключительно на данных акустической фонетики. Для каждого постулируемого признака находятся определенные акустические характеристики, относительно которых утверждается, что они однозначно определяют данный признак. Не будучи специалистом в вопросах акустической фонетики, автор не решается высказать суждение о реальных соответствиях, приводимых В. Яссемом; однако вызывает некоторое сомнение общее положение о том, что дифференциальный фонологический признак оказывается полностью выводимым из акустического анализа. Последние данные Л. В. Бондарко и Л. Р. Зиндера свидетельствуют скорее об обратном: «Между дихотомической классификацией фонем и артикуляторно-акустическими признаками нет строгой связи. Физические характеристики фонем не являются инвариантными и изменяются в зависимости от фонетической позиции» [29]. Это отсутствие строгого одно-однозначного соответствия между фонологическим признаком и акустической картиной, фиксирующей индивидуаль-

---

[28] W. J a s s e m. The distinctive features and the entropy of the Polish phoneme system, стр. 87—108.

[29] Л. В. Б о н д а р к о, Л. Р. З и н д е р. Дифференциальные признаки фонем и их физические характеристики, стр. 37. Ср. высказывание Г. Фанта на том же конгрессе: «The crucial question concerns the integrity of these (distinctive.—*Д. С.*) features on the level of the acoustic speech wave and on the level of perception. How far in abstraction does one have to go in order to formulate the common denominator of one feature in various contexts and how much can one simplify the description of the feature without losing its differentiating function? Is there a unique perceptual quality underlying each feature irrespective of context? Or is the common production model the only basis for perceptual decoding?».— «XIII Международный психологический конгресс. Москва, 1966. Симпозиум 23. Модели восприятия речи». Л., 1966, стр. 13.

Таблица 2

Матрица идентификации фонем польского языка
(по В. Яссему)

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Phoneme	*i*	*u*	*w*	*i*	*j*	*a*	*o*	*e*	*r*	*l*	*m*	*n*	*ŋ*	*ɲ*
Consonantal	—	—	—	—	—	—	—	—	+	+	+	+	+	+
Supraglottal									—	—	—	—	—	—
Nasal									—	—	+	+	+	+
Smooth									—	+				
Compact	—	—	—	—	—	+	+	+			—	—	+	+
Acute	—	—	—	+	+	—	—	+			—	+	—	+
Low-tone	—	+	+			—	+							
Short-Voiced		—	+	—	+									

15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
p	*b*	*t*	*d*	*k*	*g*	*c*	*ɡ*	*f*	*v*	*s*	*z*	*t͡s*	*d͡z*	*x*	*ɕ*	*ʑ*	*t͡ɕ*	*d͡ʑ*	*ʂ*	*ʐ*	*t͡ʂ*	*d͡ʐ*
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
—	—	—	—	—	—	—	—	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
—	—	—	—	+	+	+	+	—	—	—	—	—	—	+	+	+	+	+	+	+	+	+
—	—	+	+	—	—	+	+	—	—	+	+	+	+	+	+	+	+	+	+	+	+	+
										—	—	+	+		—	—	+	+	—	—	+	+
—	+	—	+	—	+	—	+	—	+	—	+	—	+		—	+	—	+	—	+	—	+

ное речевое событие, вовсе не подрывает дихотомическую теорию различительных признаков Р. Якобсона, Г. Фанта и М. Халле, поскольку в этой теории признаки выделяются как абстрактные различительные единицы, которые основываются не исключительно на акустических данных, но на комплексе акустических, артикуляторных и фонологических характеристик.

Впрочем, хотя В. Яссем и постулирует чисто акустическую основу для своих признаков, некоторые из них несомненно имеют артикуляторное и фонологическое обоснование. Возьмем, например, признак supraglottal—infraglottal. Этот признак четко отделяет сонанты и плавные *rlmnŋjɲ* от шумных — проблема, которая довольно трудно решается в рамках дихотомической теории, где они определяются либо как гласные и согласные, либо как

негласные и несогласные. Поэтому система В. Яссема не вызывала бы возражений, если бы подобная артикуляторная и фонологическая основа находилась бы для всех признаков. Это, однако, не так. Первое серьезное возражение вызывает трактовка польских *j* и *w* как чистых гласных, отличающихся от противопоставленных им *i* и *u* как краткие (short) от долгих (long). В то время как *r* и *l* характеризуются как согласные, *j* и *w* описываются как гласные. Однако фонологическое поведение *j* и *w* никак не может быть охарактеризовано как вокалическое: они не являются слогообразующими элементами, в то время как *i* и *u* не только являются таковыми, но составляют отдельные слоги; на *j* и *w*, следовательно, не может падать ударение. Наконец, *j* и *w* встречаются в разнообразных сочетаниях с другими гласными, в том числе с *j* (*ji*) и *u* (*wu*), в то время как подобные гомогенные сочетания запрещены для других гласных (вообще в польском языке наложены существенные ограничения на сочетания гласных друг с другом). В слоге *j* и *w* ведут себя как чистые согласные (плавные), образуя сочетания типа *jaw* или *waj*, что было бы невозможно, если бы *j* и *w* были гласными — сочетания типа VVV невозможны в польском языке. Поэтому характеризовать *j* и *w* как гласные — значит совершать насилие над фонологической системой польского языка в угоду неверно понятым акустическим данным.

Далее, характеристика отличия *j* от *i* и *w* от *u* как кратких от долгих вызывает возражения. Традиционно признак «краткость — долгота» используется для разделения фонологически кратких и долгих гласных (как, например, в чешском языке). Здесь этот признак отличает не только *j*, *w* от *i*, *u*, но и аффрикаты *c z ć ź č ž* от фрикативных *s z ś ź š ž*. Подобная интерпретация признака «краткость—долгота» делает невозможной фонологическую типологию польского языка с языками, где имеется фонологическая долгота гласных. Кроме этого, сами фонетические данные представляются нам далеко не бесспорными. В. Яссем приводит два примера: фразу On się mnie chyba boi и фразу Wdrzwiach hotelu stal portier i boj. Отмечается, что в первой фразе продолжительность сегмента [i] в последнем слове 0,25 сек., в то время как во второй — 0,16 сек. Сразу возникает вопрос, за счет чего возникает эта разница. В. Яссем молчаливо принимает как само собой разумеющееся, что это различие объясняется только различием внутренних различительных признаков фонем [i] и [j]. В то же время на уровне целого высказывания подобные (на первый взгляд не очень существенные) различия могут возникать благодаря различной ритмической структуре (в первой фразе 3 ударения, во второй — 5), длине фраз и т. д. и т. п. Кроме того, внимательное рассмотрение рис. 3—6 в статье В. Яссема показывает, что различие в длительности распространяется не только на [i]—[j], но и на предшествующие им сегменты [o]— в первой фразе его длительность примерно 0,25 сек., а во второй — 0,17 сек. Почему

признаком, различающим оба слова, было избрано различие [i]—[j], а не [o_1]—[o_2]?

Далее указывается, что различие между польскими аффрикатами и гоморганными фрикативными — только в длительности фрикативного сегмента. Возможно, что это различие действительно релевантно, однако данные, показывающие поведение обеих групп в разных фонетических позициях, не приводятся. Во всяком случае, нам представляется, что основывать выделение признака на различной фонетической (чисто физической) длительности сегментов крайне рискованно (сравним данные по изофонам количества в немецких диалектах, показывающие, что в некоторых диалектах различие в физической длительности долгих и кратких гласных почти устранено, что не мешает выделению в этих диалектах фонологической категории долготы—краткости) [30].

Поэтому нельзя принять предлагаемую В. Яссемом схему различительных признаков. Помимо того, что они оказываются отнюдь не универсальными, а основываются на акустических данных, извлеченных из анализа лишь одного языка, сами эти данные в решающих пунктах проблематичны.

Таким образом, нами принимается следующая система польских фонем: *rnńŋlmḿjtdczszpṕbb́fv́vśšžźčćžźkḱgǵxeẽoõuia*. Итого 42 фонемы, которые являются базисными единицами подсчета.

§ 2. Исходные данные. Эксперименты по проверке однородности текстов относительно частот фонем

Исходными данными любого статистического эксперимента служат наблюденные частоты основных единиц подсчета. Для того, чтобы получить эти данные, все тексты, указанные в § 1 настоящей главы, были переписаны по вышеприведенным правилам внутреннего сандхи в фонологическую транскрипцию, и была подсчитана частота каждой фонемы в каждом тексте. В число подсчитываемых единиц фонологического уровня не входит пауза, которая, однако, естественно влияет на запись текста. Все последующие цифры относятся только к реальным фонемам. Но это не все: чтобы сравнивать частоты внутри одного текста — два более длинных текста (1-я глава повести Я. Ивашкевича и 1-й акт пьесы Л. Кручковского) были разделены на более мелкие зоны: первый текст — на 5 зон, а второй — на 7 зон, и была отдельно подсчитана частота фонем в каждой из зон. Каждый рассказ С. Мрожека и Е. Шанявского также приравнивался к отдель-

---

[30] E. Zwirner. Phonometrische Isophonen der Quantität der deutschen Mundarten. «Phonetica», 1959, № 4, Supplement, стр. 93—125.

ной зоне. Итого весь просмотренный текст разделен на 22 зоны. В целом объем обследованных текстов составил [31]:

Всего — 103 828 фонем

1. Текст Я. Ивашкевича — 15 319 фонем
1 зона — 3148
2 зона — 3149
3 зона — 3105
4 зона — 3151
5 зона — 2766

2. Текст Л. Кручковского — 26 589 фонем
1 зона — 3419
2 зона — 3584
3 зона — 3718
4 зона — 3656
5 зона — 3567
6 зона — 3659
7 зона — 4986

3. Рассказы Е. Шанявского — 27 314 фонем
Profesor Tutka o złodzieju — 3809
O slowie drukowanym — 5196
Sprawa osobista — 5683
O pszczołach i miodzie — 6212
O dwóch malowidłach — 6414

4. Рассказы С. Мрожека — 34 606 фонем
Góral —4874
Interwał — 6427
Nadzieja — 6931
Mały przyjaciel — 7559
Jak walczyłem — 8815

В приложении (стр. 000) содержатся все данные об абсолютных и относительных частотах фонем в каждой из 22 зон.

Размер минимальной выборки (одна зона) колеблется в нашем эксперименте от 2766 до 8815. Насколько такой объем является оптимальным? По-видимому, при проверке текста на однородность вполне естественным будет такое требование к выборке, чтобы она по крайней мере обеспечивала однократную встречаемость любой из 42 фонем. Практически задача сводится к обеспечению хотя бы однократной встречаемости наиболее редкой фонемы. Такой фонемой в польском языке является ʒ. В нашем эксперименте она встречается всего 18 раз на 103 828 фонем, т. е. ее относительная частота равна примерно 1/5768.

Вероятность того, что данный объект встретится не менее 1 раза, равна $Q = 1 - (1 - p)^N$, где Q — заданный порог вероятности, p — частота объекта, а N — объем выборки. Положим эту вероятность равной 0,5. Тогда получаем:

$$0{,}5 = 1 - \left(1 - \frac{1}{5768}\right)^N, \text{ или } 1 - \left(\frac{5767}{5768}\right)^N = 0{,}5, \text{ или}$$

$$\left(\frac{5767}{5768}\right)^N = 0{,}5, \text{ тогда } N = \frac{\log 0{,}5}{\log 5767 - \log 5768} =$$

$$= \frac{0{,}30100}{0{,}00008} \approx 3760.$$

[31] В настоящем изложении цифровые данные окончательно уточнены по сравнению с материалами, опубликованными в нашей статье «Статистическая однородность текста на фонологическом уровне в польском языке»

Таким образом, объем наших выборок в общем соответствует искомому. Однако следует отметить, что фонема ʒ по своей частоте резко выделяется даже среди других редких фонем: чаще ʒ зафиксирована лишь фонема ē, но ее частота (92) в пять раз выше частоты ʒ.

Мы уже отмечали, что выборка должна быть не ниже некоторой минимальной величины, чтобы снизить возможные неоднородности за счет непопадания в выборки определенных фонем. В самом начале эксперимента эти соображения учитывались нами, однако определенные технические ограничения заставляли придерживаться определенного объема выборки (немногим меньше 3 600). Дело в том, что частоты фонем в первом тексте (Я. Ивашкевич) подсчитывались с помощью электронно-счетной машины «Урал-1» в Киевском государственном университете [32]. Максимальный объем текста, обрабатываемого за один раз на этой машине, составляет 3 600 единиц (при машинном подсчете учитывались знаки паузы между фонетическими словами). Соответственно объем первых четырех зон текста Я. Ивашкевича равен 3 600 — n, где n — количество пауз в данной зоне. Объем пятой зоны представляет собой остаток.

Текст Л. Кручковского делился на зоны, немного большие, чем текст Я. Ивашкевича, но примерно такого же порядка (3000—3 600). Объемы рассказов С. Мрожека и Е. Шанявского были заданы естественным образом. Разумеется, что, поскольку мы имеем дело со случайным процессом, объем текста примерно в 3 700—3 800 фонем вовсе не обеспечивает автоматической встречаемости фонемы ʒ. Например, в самой длинной выборке в 8 815 фонем ʒ не встретилось ни разу, а в выборке в 3 148 фонем ʒ встретилось два раза. Всего в половине наших выборок встречаются 42 фонемы и в половине — встречается по 41 фонеме (в 9 случаях отсутствует ʒ, а в двух ē). Подобные данные можно считать вполне удовлетворительными, поскольку в подсчетах, приводимых в литературе (например, у Г. Хердана), в различных выборках, как правило, могут отсутствовать по 4—5 редких фонем.

Таким образом, частотные ряды в каждой выборке, полученные нами в результате подсчетов, являются экспериментальными репрезентантами частотного распределения всего набора фонем, а не лишь некоторых из них; соответственно эти частотные ряды можно сравнивать между собой: частоте каждого члена обнаруживается необходимое соответствие.

(«Структурная типология языков». М., 1966), где объем всех текстов был указан 103 815, а не 103 828, как следует.

[32] Автор пользуется случаем выразить самую искреннюю благодарность Л. С. Стойковой и Н. С. Горбатюк, производившим машинный подсчет фонем в первом тексте.

* * *

Для сравнения между собою полученных частотных рядов необходимо принять определенную начальную гипотезу (в статистике ее называют нулевой гипотезой H_0). В качестве H_0 всегда выступает предположение о совпадении вероятностей (функций распределения, законов распределения). Это предположение затем проверяется с помощью статистических критериев.

В качестве статистического критерия нами был избран так называемый критерий χ^2. В статистике существует несколько видов критерия χ^2, применяющихся для сравнения наблюденной вероятности и теоретической, для сравнения двух наблюденных вероятностей, для сравнения несколько наблюденных вероятностей, для проверки нормальности распределения и т. п.

Общий смысл критерия χ^2 в применении к языковым данным сводится к тому, что вычисляется разность между наблюденными частотами одной и той же величины; эта разность затем возводится в квадрат и определенным образом преобразуется так, что получившаяся в результате преобразования величина (ее называют χ^2) может быть сравнена с некоторым эталоном. Если в результате сравнения оказывается, что эта величина, репрезентирующая различие частот, больше или равна эталону, то H_0 отвергается; если же она оказывается меньше эталона, то H_0 не отвергается. Подобное сравнение оказывается возможным, так как величина, получающаяся в результате преобразования разности частот, подчиняется так называемому распределению χ^2 (поэтому критерий и называется χ^2)[33], а эталонами служат заранее вычисленные значения распределения χ^2, содержащиеся в специальных таблицах[34].

Почему для сравнения частотных рядов был избран именно критерий χ^2? В статистике существуют и другие методы оценок наблюденных частот, например, вычисление доверительных границ для вероятности (т. е. верхней и нижней границы, выход за пределы которых крайне маловероятен), критерий Стьюдента, позволяющий определить, принимает ли вероятность некоторое гипотетическое среднее значение, и т. п. Эти методы, однако, основываются на допущении, что все наблюденные случайные величины независимы и одинаково нормально распределены.

Поскольку для лингвистических частот предположение о нормальном распределении экспериментально не проверялось, применение критериев, основанных на нормальности, оказывается невозможным. Об этом, между прочим, пишет Р. Абернати в своей рецензии на книгу С. С. Ахмановой, И. А. Мельчука, Е. В. Падучевой и Р. М. Фрумкиной «О точных методах исследования языка»:

[33] О распределении χ^2 см. стр. 118 и далее в уже цитированной книге Б. Л. ван дер Вардена «Математическая статистика».

[34] Там же. Приложение.

«...имеются гораздо более удобные критерии, например, χ^2, которые могут оказаться весьма полезными в тех случаях, когда априори нельзя сделать определенных предположений о форме распределения» [35].

Применение более грубых методов, например сравнение стандартных ошибок двух рядов частот, представляется неоправданным, во-первых, поскольку при этих методах распределения представляются весьма суммарно и нельзя сделать выводов о равенстве (или неравенстве) отдельных частот и, во-вторых, потому, что предполагается, что частоты всех фонем (42 в нашем случае) обязательно стремятся к одной и той же средней величине, т. е. представляют одну вероятность, что явно не так.

Таким образом, критерий χ^2 был избран, во-первых, как позволяющий сравнивать частотные ряды с неизвестным законом распределения и, во-вторых, как позволяющий устанавливать достаточную (или недостаточную) близость конкретных сравниваемых значений. Это свойство критерия следует подчеркнуть особо. Неотвержение H_0 в случае критерия χ^2 означает неотвержение гипотезы о фактическом равенстве соответствующих конкретных частот в сравниваемых рядах. То, что с помощью критерия χ^2 подвергается проверке равенство частот соответствующих фонем, будет видно из примеров, которыми мы в дальнейшем проиллюстрируем применение критерия: вычисляется χ^2 для каждой фонемы, поэтому достаточно сильное расхождение частот всего нескольких фонем из 42 может привести к образованию слишком большого значения накопленного χ^2. Таким образом, критерий χ^2 является весьма сильным, иными словами, обращаясь к критерию χ^2, мы должны формулировать довольно сильную нулевую гипотезу H_0: сравниваемые частотные ряды представляют одно распределение и соответствующие частоты практически равны.

Напомним, что общая форма критерия χ^2 для проверки гипотезы вероятности равна

$$\chi^2 = \sum \frac{(x_i - np_i)^2}{np_i},$$

где x_i — наблюденные количества выборочных элементов, принадлежащих первому из m классов, на которые разбита выборка объема n из некоторой бесконечной совокупности, а p_i — теоретические вероятности, являющиеся определенными заранее заданными числами.

В нашем эксперименте не делается никаких предположений о теоретической вероятности. Следовательно, требуется сравнить друг с другом экспериментально найденные ряды частот; для подобных случаев критерий χ^2 приобретает специальную форму.

[35] R. Abernathy. Review. «International Journal of American Linguistics», v. 33, 1967, № 1, стр. 87.

Прежде чем перейти к изложению экспериментальных результатов, поясним, каким образом ведется работа с критерием χ^2. В любом пособии по математической статистике [36] можно найти таблицу критических знаний распределения χ^2. Вот как выглядит одна строка из этой таблицы:

f/q	99,95	99,90	99,5	99,0	97,5	95,0	90,0	80,0	70,0	60,0
41	17,5	18,6	21,4	22,9	25,2	27,3	29,9	33,3	35,8	38,1

f/q	50,0	40,0	30,0	20,0	10,0	5,0	2,5	1,0	0,5	0,1	0,05
41	40,3	42,7	45,2	48,4	52,9	56,9	60,6	65,0	68,1	74,7	77,5

Символ f обозначает число, известное под названием к о л и ч е с т в о с т е п е н е й с в о б о д ы. Практически оно участвует в критерии, обозначая (не совсем прямым образом) количество сравниваемых величин, или количество выборок, в которых рассматривается одна величина, или количество величин и количество выборок. Таким образом, задается число степеней свободы, которое может принимать данная переменная.

В формуле критерия, приведенной выше, f устанавливается равным m — 1 (т. е. количество сравниваемых величин минус 1). Для каждой формулы χ^2 количество степеней свободы определяется по-своему.

Мы специально привели здесь строку с f = 41, поскольку в нашем случае количество сравниваемых величин (частот фонем) равно 42, следовательно, число степеней свободы будет равно 41. Обычно в таких таблицах степени свободы даны в вертикальном столбце от 1 до 100. Следовательно, эта таблица рассчитана максимум на 100 состояний переменной. В горизонтальной строчке мы имеем значения q, так называемых процентных точек распределения χ^2, от 99,95% до 0,05%. На том, что представляет собой число q в содержательном плане, следует остановиться подробнее.

Практически, когда вычислено экспериментальное χ^2, по заданному извне числу степеней свободы в таблице находят наименьшее значение табличного χ_q^2, превосходящее экспериментальное χ^2. Предположим, что значение χ^2 при q = 41 достигло 50. Ищем, каково то наименьшее χ_q^2, которое больше 50, и обнаруживаем в нашей таблице, что это 52,9. Этому значению χ_q^2 соответствует q = 10%, именно по этой величине q и определяется, насколько

[36] См., например, табл. 6 в книге ван дер Вардена, а также таблицы в книгах: Я. Я н к о. Математико-статистические таблицы. М., 1961; и Л. Н. Б о л ь ш е в а и Н. В. С м и р н о в а. Таблицы математической статистики. М., 1966.

экспериментальное χ^2 удовлетворяет нулевой гипотезе H_0. Величина q в критериях проверки называется уровнем значимости. Грубо говоря, для того, чтобы не отвергнуть H_0, q должно быть достаточно велико.

К сожалению, понятие уровня значимости относится к числу теоретически наименее определенных. Более того, существуют даже расхождения в том, что считать уровнем значимости. Поэтому нам придется, оговорив свою недостаточную компетентность, остановиться на этом понятии и попытаться эксплицировать его смысл.

Величина уровня значимости соответствует вероятности того, что в нашем опыте будут зафиксированы события, которые мы по чисто семантическим соображениям считаем практически неосуществимыми. Обычно в приложениях статистики задают уровни значимости, равные 10, 5, 2, 1%. Это значит, что по определенным внешним причинам мы принимаем, что в десяти, пяти, двух или одном случае из ста может произойти событие, которое практически неосуществимо. После того, как выбран уровень значимости q = 10, 5, 2 или 1% и т. п., определяется критическая область данного критерия, вероятность попадания в которую в случае, когда гипотеза верна, в точности равна уровню значимости. Если экспериментальное значение критерия, вычисленное по наблюденным частотам, окажется в критической области, мы бракуем нулевую гипотезу.

Итак, в приведенной выше строке все значения критерия, начиная с 52,9 (соответствующего q = 10% в предположении, что был избран именно этот уровень значимости), т. е. 52,9; 56,9; 60,6; 65,0; 68,1; 74,7; 77,5, лежат внутри критической области. Таким образом, фактически лишь крайняя правая часть таблицы оказывается существенной для проверки по критерию χ^2. Очень часто (например, в цитированной книге Б. Л. ван дер Вардена) таблица критических значений χ^2 включает значения, соответствующие лишь q = 0,1; 0,05 и 0,01, а все предыдущие значения в таблице не приводятся, поскольку они лежат вне критической области и не нужны для практической работы с критерием. Значения критерия, лежащие вне критической области, образуют дополнительную к ней область допустимых значений.

Если q (в десятичных дробях) — уровень значимости, то вероятность попадания критерия в область допустимых значений при справедливости данной гипотезы равна 1 — q и соответственно (для вышеприведенных значений q) 90, 95, 98 или 99%. Если значение критерия окажется в области допустимых значений, то мы еще не можем утверждать, что нулевая гипотеза подтвердилась. Мы можем только заключить, что наблюденное значение критерия не противоречит ей.

Последнее замечание весьма существенно. Оно должно предостеречь против неверной интерпретации статистических резуль-

татов. С помощью критерия проверки мы можем либо о т в е р г н у т ь гипотезу, либо н е отвергнуть ее, но решительно п р и н я т ь гипотезу с помощью лишь статистических данных н е в о з м о ж н о.

Вопрос о семантическом содержании уровня значимости важен хотя бы потому, то не существует никаких статистических обоснований выбора того или иного уровня значимости. В учебниках и пособиях обычно ссылаются на практику применения критериев, на опыт, на обстановку исследования и т. п. Например, Е. Вентцель подчеркивает в своей книге «Теория вероятностей», что уровень значимости выбирается исключительно из соображений практического удобства. Однако нигде мы не нашли более подробных соображений на этот счет — в книге Б. Л. ван дер Вардена уровни значимости вводятся чисто эвристически — указывается, что можно выбрать либо 0,1, либо 0,05, либо 0,01.

Нам представляется, что нижеследующие рассуждения могут оказаться полезными. Предположим, что мы устанавливаем уровень значимости критерия, равный 0. Соответственно критическая область критерия оказывается равной 0, а вероятность попасть в область допустимых значений — 1. В этом случае любые значения случайной величины допустимы, и нет таких, которые мы могли бы считать невозможными. Вместе с тем при данном уровне значимости мы действуем абсолютно без всякого риска. Для нашего конкретного случая это означает, что частоты фонем могут как угодно отличаться друг от друга, и нулевая гипотеза все равно будет верной; но нулевая гипотеза состоит как раз в том, что эти частоты предполагаются не отличающимися. Следовательно, при $q = 0$ неотвержение нулевой гипотезы приводит к противоречию с реальной ситуацией. Это и соответствует утверждению, что уровень значимости (говоря иначе, практический смысл) критерия равен нулю, т. е. результат не имеет смысла.

Соответственно, чем выше уровень значимости, тем больше значимость результатов. С повышением уровня значимости мы сужаем круг возможных значений случайной величины. В нашем случае это означает сужение возможных колебаний частоты фонемы. Но что произойдет, если мы сильно повысим уровень значимости,— скажем, до 0,99? Тем самым мы придем к утверждению, что практически невозможными являются 99% всех наблюденных случаев и область допустимых значений составит лишь 1%. Заметим попутно, что если в эксперименте мы бы задались уровнем значимости 0,99, это бы значило, что мы требуем от наблюденных значений, чтобы они практически абсолютно совпадали, чего в опыте никогда не будет.

Если же мы положим уровень значимости равным единице, то тем самым столь же нарушим содержательность эксперимента, как и при $q = 0$. При $q = 1$ мы достигаем абсолютной значимости результатов, но ценой признания их практически неосуществи-

мыми. В самом деле, критерий проверки теряет всякий смысл, если сравниваемые величины заранее абсолютно тождественны.

Итак, выбирая уровень значимости, мы должны осознавать, что чем более высокую величину мы задаем, тем большие ограничения мы накладываем на возможные колебания частот, тем большей точности мы требуем от данных и, наоборот, чем ниже уровень значимости, тем большие колебания мы допускаем, тем меньшей точностью будут отличаться данные.

Реально мы видим, что самый высокий уровень значимости, принятый в статистической практике,— это 0,1 (10%). Гораздо более приняты уровни 0,05 и 0,01.

Таким образом, на практике статистика допускает довольно значительные колебания наблюденных частот. Если эти колебания дают величину критерия, лежащую в области допустимых значений, то говорят, что наблюденные расхождения объясняются случайностью; если же эта величина лежит в критической области, то говорят, что расхождения настолько велики, что не могут объясняться случайностью и что, следовательно, указанные значения принадлежат различным распределениям, т. е. отвергается H_0 о совпадении распределений.

Какой уровень значимости следует принять в статистическом исследовании фонологических данных с помощью критерия χ^2? Вот что пишет по этому поводу Г. Хердан: «Уровень значимости, равный 0,05, широко используется в медицинской статистике, а также в экономической статистике не вследствие каких-либо позитивных теоретических соображений, а исключительно в силу практических условий эксперимента, связанных с предметом приложения статистики. В других областях, например в статистической физике, применяется уровень значимости $q = 0{,}01$ или $q = 0{,}003$. Последний соответствует отклонению от истинной вероятности на $\pm 3\sigma$, что считается допустимым. Я уже указывал, что в лингвистической статистике более низкие уровни значимости следует считать истинным водоразделом между значимостью и незначимостью критерия» [37].

К сожалению, Г. Хердан подробнее не аргументирует свое мнение. Нам представляется, что для обоснования выбора более низких значений уровня значимости в лингвостатистике могут оказаться существенными следующие соображения. Статистическая физика имеет дело с более редкими событиями, чем медицинская или экономическая статистика. Соответственно уровень значимости снижается, поскольку редкие события встречаются лишь в большом количестве опытов — предположим, один раз на 1000 опытов. Если при этом допустить, что уровень значимости равен 5%, это будет означать вероятность совершения пяти прак-

[37] G. Herdan. The Advanced Theory of Language as Choice and Chance, стр. 415.

тически неосуществимых событий на 100 опытов, т. е. 50 на 1000. Таким образом оказывается, что количество событий, которые мы полагаем практически невозможными, больше, чем количество наблюденных положительных исходов опыта, а это явно неудобно. Соответственно, чтобы привести условия применения критерия в соответствие с реальными данными эксперимента, требуется понизить уровень значимости (до 0,0005). Это снижение в случае лингвистических объектов фонологического уровня объясняется аналогичным образом: частота некоторых из этих объектов довольно мала, поскольку мы имеем дело с их одно- или двукратной встречаемостью в выборках весьма большого объема (порядка десятков тысяч элементов). Поэтому и в случае лингвистических объектов разумно снижение уровня значимости до 1 события на 1 000 или ниже. Напомним, что в статистической психологии применяется (по сходным причинам) уровень значимости $q = 0{,}0005$. Таким образом, при использовании критерия χ^2 для сравнения частот фонем в различных текстах целесообразно избрать более низкий уровень значимости, чем уровень значимости $q = 0{,}05$, обычно принятый во всех учебниках по математической статистике.

Практически мы каждый раз будем указывать величину q_i, которой соответствует найденное χ^2, и в зависимости от этой величины будет ясно, какой уровень значимости критерия можно избрать.

Итак, познакомившись с основными принципами применения критерия χ^2 и с понятиями числа степеней свободы и уровня значимости, приступим к изложению практических результатов.

* * *

1. Прежде всего мы решили применить критерий χ^2 для проверки стабильности частот фонем в двух произвольно взятых выборках. Как мы уже отмечали, в этом тесте ряд из 42 чисел, полученных из одной выборки, сравниваем с аналогичным рядом, полученным из другой выборки.

Выборка I (объем n)	Выборка II (объем m)
$\nu_1(\alpha)$	$\mu_1(\alpha)$
$\nu_2(\beta)$	$\mu_2(\beta)$
$\nu_3(\gamma)$	$\mu_3(\gamma)$
.	.
.	.
.	.
$\nu_{42}(\varepsilon)$	$\mu_{42}(\varepsilon)$

где ν_i — абсолютные частоты фонем (α, β, γ . . . ε) в выборке I объемом n элементов, а μ_i — абсолютные частоты тех же фонем (α, β, γ, . . . ε) в выборке II объемом m элементов.

Для этого случая двух выборок критерий χ^2 имеет следующую форму:

$$\chi^2 = mn \sum_1^u \frac{1}{\mu_i + \nu_i} \left(\frac{\mu_i}{m} - \frac{\nu_i}{n} \right)^2 .$$

Количество степеней свободы равно (k —1) = 42 —1 = 41. Очевидно, что $\frac{\mu_i}{m}$ и $\frac{\nu_i}{n}$ — относительные частоты соответствующих фонем. Член $\frac{1}{\mu_i + \nu_i}$ вводится для приведения разницы между относительными частотами в соответствие с объемом выборок. Для каждой фонемы вычисляется выражение $\frac{1}{\mu_i + \nu_i} \left(\frac{\mu_i}{m} - \frac{\nu_i}{n} \right)^2$, составляющее индивидуальное значение χ^2 для каждого сравниваемого разряда (т. е. фонемы). Затем все 42 индивидуальных значения χ^2 суммируются, давая то же χ^2, но уже для всего ряда фонем. Затем для приведения этого значения в соответствие с различными объемами выборок $\sum_1^u \chi^2$ умножается на произведение mn, давая истинное значение наблюденного χ^2, которое затем проверяется в таблице критических значений χ_q^2 в строчке f = 41. Если сравниваются выборки одинакового объема, выражение принимает более простой вид:

$$\chi^2 = \sum_1^u \frac{(\mu_i - \nu_i)^2}{\mu_i + \nu_i} .$$

В самом начале нашего эксперимента мы не имели никаких предположений о том, какого рода результаты следует ожидать. Было явно бессмысленным вычислять критерий χ^2 для проверки однородности всех возможных парных сочетаний из имеющихся 22 выборок — это было бы слишком трудоемко и нерационально. Единственным разумным методом было вести вычисления до тех пор, пока в нашем распоряжении не окажутся результаты противоположного характера, т. е. пока критерий χ^2 не покажет о д н о р о д н о с т ь двух выборок по отношению к частотному распределению фонем н а р я д у с н е о д н о р о д н о с т ь ю для других двух выборок. Предполагалось, что если критерий устойчиво будет показывать только однородность или только неоднородность, то будут привлечены контрольные выборки из другого материала, и если они подтвердят результаты на наших выборках, то можно будет с определенной вероятностью сделать вывод о том, что выборки из связных текстов (соответственно сами связные тексты) однородны (или неоднородны) относительно частотного ряда 42 фонем польского языка.

Реальные эксперименты четко показали, что подобного рода устойчивые результаты при применении критерия χ^2 для проверки однородности двух произвольных выборок относительно частотного распределения всего ряда фонем на практике не встречаются.

Мы начали проверку с текста повести Я. Ивашкевича «Девушка и голуби». Были взяты первые четыре зоны этого текста, поскольку их объемы весьма близки (3148, 3149, 3105 и 3151). Приведем для примера одну таблицу вычисления критерия χ^2 (между 2-й и 4-й зонами). В дальнейшем мы не будем приводить подобных вычислений, поскольку они носят вспомогательный характер, но здесь при первом знакомстве с этим критерием нам представляется необходимым познакомить читателя с реальной практикой вычисления (см. табл. 3).

Наименьшее значение табличной величины χ^2_q, превосходящее 35,34 при К = 41, равно 35,8. Этому значению χ^2_q соответствует $q_i = 0{,}7$ (q_i — значение q, найденное в эксперименте), что заведомо гораздо выше любого стандартного уровня значимости. Следовательно, значение критерия χ^2, вычисленное в эксперименте, находится в области допустимых значений, следовательно, нулевая гипотеза о совпадении частотных распределений в двух выборках и о совпадении конкретных частот фонем не отвергается. Рассмотрение отдельных значений χ_i^2 показывает, что вывод Г. Хердана о том, что основная часть накопленного χ^2 образуется за счет наиболее редких фонем, неверен и объясняется недостаточными размерами выборок в его эксперименте. В нашем случае наибольшие χ_i^2 дают фонемы *b* (2,27), *z* (3,70), *ś* (5,44), *v* (3,55), *ź* (3,60) и *x* (3,95). Из них только *ź* является редкой фонемой, а все остальные фонемы имеют среднюю частоту. Таким образом, устойчивость (или неустойчивость) частоты фонемы нельзя непосредственно и однозначно связать с малой величиной этой частоты.

Суммарные результаты по применению критерия χ^2 для проверки однородности частот всего ряда фонем в двух произвольных выборках из текста Я. Ивашкевича выглядят следующим образом:

Зоны	χ^2	q_i	Зоны	χ^2	q_i
1—2	50,7	0,15	2—3	55,95	0,05
1—3	79,65	0	2—4	35,34	0,70
1—4	48,6	0,19	3—4	46,75	0,24

Как видим, не понадобилось даже выходить за пределы одного текста, чтобы получить противоположные результаты — опровержение нулевой гипотезы для зон 1—3 и весьма высокое значение q_i (0,70) для зон 2—4.

Таблица 3

Фонемы / Абс. частота в зонах	*p*	*b*	*m*	*ṕ*	*ź*	*b́*	*a*	*t*	*d*
2	86	36	85	5	39	19	344	126	77
4	93	50	77	5	45	14	327	144	79
χ_i^2	0,27	2,27	0,40	0	0,43	0,79	0,43	1,20	0,02

n	*ć*	*š*	*ʒ́*	*e*	*s*	*z*	*ḿ*	*ś*	*c*	*v́*
126	38	63	22	334	88	64	31	56	51	34
138	30	55	25	308	87	44	23	56	30	26
0,54	0,94	0,54	0,19	1,05	0	3,70	1,18	0	5,44	1,07

o	*f*	*v*	*ń*	*f́*	*č*	*ź*	*i*	*k*	*g*	*µ*
260	41	93	76	3	32	2	263	87	54	104
283	43	69	83	3	37	8	286	77	57	104
0,97	0,05	3,55	0,31	0	0,36	3,60	0,96	0,61	0,08	0

ḱ	*ʒ́*	*ǵ*	*u*	*x*	*r*	*l*	*õ*	*ẽ*	*ǯ*	*j*
28	6	5	89	21	97	57	20	3	1	89
31	11	4	101	36	101	70	17	3	0	73
0,15	1,47	0,11	0,76	3,95	0,08	1,33	0,29	0	1	0,77

$\Sigma\chi_i^2 = 35{,}34$.

Эти результаты показывают, что не удается сделать однозначного вывода об однородности или неоднородности двух произвольно взятых выборок даже из одного текста относительно частот всего ряда фонем. Можно лишь отметить, что окончательно негативный результат получился в одном случае из шести. Иными словами, если взять две произвольные выборки из одного текста, то скорее можно ожидать, что нулевая гипотеза при проверке критерия χ^2 не будет отвергнута. Существенно, однако, то, что

эта возможность чисто статистическая. Ожидать опровержения гипотезы в некоторых случаях вполне можно (особенно когда сравниваются отделенные друг от друга части текста).

Каков должен быть следующий шаг нашего эксперимента? Должны ли мы проверять попарно для каждой выборки ее однородность относительно другой? Разумеется, нет: нам было важно установить в принципе однородность или неоднородность двух произвольных выборок. Поскольку полученные результаты относятся к одному связному тексту, было решено сравнить между собой несколько связных текстов. Вот наши результаты:

	χ	q_i
С. Мрожек. «Góral» — «Interwał»	66,419	0,007
Е. Шанявский. «Profesor Tutka o złodzieju», «O słowie drukowanym»	74,493	0,001

Очевидно, что при переходе от выборок из одного текста к выборкам, представляющим собою различные тексты, шансы на однородность падают. В принципе можно, наверное, считать, что полученные q_i лежат в области допустимых значений, хотя они весьма малы; даже если сделать такое допущение, то все равно различие между значениями q_i для выборок из одного текста и в случае различных текстов сразу бросается в глаза. Тенденция — к явному снижению возможности однородности по мере того как тексты становятся замкнутыми.

Далее, перейдем от отдельных рассказов к более крупным текстам и сравним между собою весь текст Я. Ивашкевича и весь текст Л. Кручковского. Получаем $\chi^2 = 401{,}904$ и $q_i = 0$. Тенденция совершено очевидным образом продолжается — если брать тексты не только различные и связные, но и достаточно большие по объему, то гипотеза об однородности должна быть отвергнута. Так же обстоит дело и в случае сравнения двух текстов, каждый из которых является объединением более мелких связных текстов:

		χ^2	q_i
Пять рассказов Е. Шанявского	— Пять рассказов С. Мрожека	125,609	0

Наконец, для подтверждения результатов текст Я. Ивашкевича был сопоставлен с текстом философской статьи А. Шаффа «Marksizm a filozofia człowieka». Получен вполне предсказуемый результат $\chi^2 = 576{,}78$, $q_i = 0$.

Можно было бы, наверное, сравнивать между собою и другие крупные тексты, но, по-видимому, ничего нового к полученным нами результатам не прибавится.

Таким образом, после первого эксперимента можно сделать следующие выводы.

Постулат Г. Хердана об обязательной однородности текстов относительно частот всего ряда фонем при проверке по критерию χ^2 (а именно этим критерием Г. Хердан советует пользоваться) должен быть отвергнут.

Однородность может не отвергаться для выборок из одного текста, однако и в этом случае возможность неоднородности вполне реальна. Таким образом, отдельный связный текст нельзя считать случайной выборкой из генеральной общеязыковой совокупности относительно частот всего ряда фонем. Такой текст скорее сам может представлять собою некоторого рода генеральную совокупность относительно частот, полученных из частей этого текста. Общая тенденция такова, что шансы на однородность уменьшаются при переходе от соседних частей одного текста к частям, расположенным на расстоянии; еще более возможность однородности уменьшается при переходе к сравнению небольших связных текстов и, наконец, эта возможность становится нулевой при сравнении сколько-нибудь крупных связных текстов.

По нашему мнению, полученные нами отрицательные результаты достаточно красноречиво опровергают суждение Г. Хердана о том, что частоты фонологического уровня в принципе не зависят от вида текста и выборки из него. Наоборот, явственно продемонстрировано, что эти частоты реально зависят от таких характеристик текста, как его законченность, объем, тема.

В этом плане частоты фонем оказываются подобными частотам слов. Этот вывод об определенной изоморфности статистического строения словаря и инвентаря фонем будет подкреплен некоторыми последующими данными. Здесь же отметим, что подобная зависимость частот фонем от факторов организации текста впервые выявлена статистически в настоящей работе. При этом специфика статистических доказательств такова, что немногих отрицательных результатов оказывается вполне достаточно, чтобы опровергнуть построения, основанные на гораздо большем количестве положительных результатов. Поскольку положение о независимости частот фонем от вида организации текста было основано (как мы пытались показать) на весьма скромном числе положительных результатов, мы можем утверждать, что результаты, полученные в настоящей работе, могут претендовать на определенную дефинитивность.

Не забудем, однако, что речь идет о весьма сильном критерии, в основу которого кладется гипотеза о том, что наблюденные частоты каждой фонемы в принципе можно считать равными, поскольку различия в частотах объясняются только случайностью. Итогом первого эксперимента явился результат, который в определенном смысле можно считать отрицательным. Вследствие этого теряет смысл проведение сравнения других двух произвольных выборок. Следует отметить, что в принципе определение однородности при сравнении двух произвольных выборок является технически и

теоретически довольно трудной задачей: что такое произвольные выборки (в статистике, как правило, не дается экспликация этого понятия; естественно, что в лингвистической статистике под произвольными обычно понимаются любые выборки), где предел числа сравнений, за которым любые две выборки для данного языка однородны относительно частот всего ряда фонем? Поэтому следует рассмотреть структуру однородности и неоднородности текстов относительно частот фонем уже с других точек зрения. Можно было бы одновременно сравнить частоты фонем в каждом из текстов и во всех текстах сразу с помощью другой разновидности критерия χ^2 — многомерного критерия χ^2. Однако это не представляется ни возможным (в частности, при вычислении этого критерия для всех 22 зон необходимо произвести суммирование достаточно сложно вычисляемых выражений по $42 \times 22 = 924$[!] разрядам, а для этого уже нужна электронно-счетная машина), ни целесообразным, поскольку доказана возможность неоднородности при попарном сравнении текстов, а при возрастании числа сравниваемых членов шансы на однородность сильно уменьшаются.

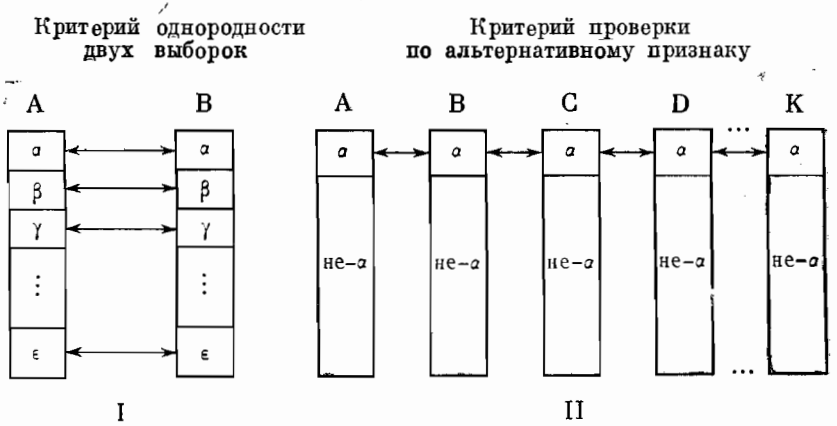
2. Поэтому для следующего шага эксперимента нами был избран другой вид критерия χ^2, а именно так называемый критерий проверки по альтернативному признаку[38]. Этот критерий предусматривает, что вся выборка делится на две части: содержащие признак А и не содержащие его. В нашем случае вся выборка может делиться, например, на следущие части — одна, в которую входят члены, обладающие признаком «быть фонемой *t* (или *k*, *a* и т. п.)», и вторая альтернативная первой, куда входят члены, не обладающие указанным признаком. Если имеется k выборок (в нашем случае сравниваемых зон текстов) с альтернативным признаком А, который характеризуется параметрами a_1, a_2, a_3, . . . , a_k (частоты данной фонемы в различных выборках, от 1-й до k-ой), и мы должны проверить нулевую гипотезу о том, что $a_1 = a_2 = a_3 = \ldots a_k$ на уровне значимости q, то подсчитывается критерий следующего вида:

$$\chi^2 = \frac{1}{\bar{p}(1-\bar{p})} \sum_{i=1}^{k} n_i (p_i - \bar{p})^2 ,$$

где n_1 — численность выборки; a_i — количество элементов с признаком A в выборке, $p_i = \frac{a_i}{n_i}$ (т. е. относительная частота данной фонемы в данной выборке), а $\bar{p} = \frac{\sum_{i=1}^{k} a_i}{\sum_{i=1}^{k} n_i}$, т. е. средняя относитель-

[38] См.: Я. Я н к о. Математико-статистические таблицы. М., 1961.

ная частота фонемы по k выборкам. Различие предыдущего критерия с критерием проверки по альтернативному признаку можно схематически представить следующим образом:



В случае I сравниваемыми разрядами являются частоты каждой из 42 фонем, следовательно, количество сравниваемых разрядов равно количеству фонем (число степеней свободы равно числу сравниваемых разрядов минус 1), а в случае II сравниваемыми разрядами являются частоты одной и той же фонемы в каждой из K выборок, следовательно, количество сравниваемых разрядов равно количеству сравниваемых выборок (число степеней свободы равно числу сравниваемых разрядов минус 1).

Таким образом, используя критерий χ^2 для проверки по альтернативному признаку, мы можем привлекать для сравнения много выборок, а не ограничиваться двумя, как это было в первом эксперименте. При этом мы получаем возможность исследовать однородность текста по отношению к частоте всего одной фонемы, выясняя таким образом, какие фонемы в отдельности имеют более стабильную частоту, а какие — менее стабильную. Такая возможность в принципе была в первом эксперименте (поскольку вычислялись χ_i^2 для каждой фонемы), однако там мы были ограничены тем, что сравнивались всего две выборки.

Был вычислен критерий χ^2 проверки по альтернативному признаку для частоты каждой фонемы сначала внутри каждого из четырех больших текстов, а потом для всех 22 выборок вместе. Эти результаты приводятся в табл. 4.

Как и следовало ожидать, проверка по критерию χ^2 для альтернативного признака показала, что среди фонем наблюдаются как случаи стабильной частоты, так и случаи сильных колебаний. Эти противоположные случаи чрезвычайно трудно свести к единой формуле, так как для каждого из четырех рассмотренных текстов выделяются свои фонемы со стабильной частотой, не совпадающие

Однородность четырех текстов польского языка в отношении частот фонем

Фонемы	Я. Ивашкевич «Девушка и голуби» (5 выборок) Σ = 15 319			Л. Кручковский «Первый день свободы» (7 выборок) Σ = 25 689			С. Мрожек Пять рассказов (5 выборок) Σ = 34 606			Е. Шанявский Пять рассказов (5 выборок) Σ = 27 314			Сумма четырех текстов (22 выборки) Σ = 103 828		
	f	p	q_i	f	p	q_i	f	p	q_i	f	p	q_i	f	p	q_i
s	453	0,030	0,78	710	0,027	0,82	981	0,028	0,55	806	0,030	0,76	2950	0,029	0,85
ẽ	16	0,001	1	19	0,001	0,33	40	0,001	0,13	17	0,001	0,05	92	0,001	0,35
ǵ	17	0,001	0,50	13	0,001	0,72	32	0,001	0,10	21	0,001	0,025	93	0,001	0,25
ḿ	127	0,008	0,41	230	0,009	0,975	243	0,007	0,05	205	0,008	0,17	805	0,008	0,24
o	1369	0,090	0,26	2238	0,084	0,45	3112	0,090	0,28	2394	0,088	0,67	9113	0,088	0,20
u	491	0,032	0,22	827	0,031	0,45	1077	0,031	0,88	912	0,034	0,025	3307	0,032	0,20
z	265	0,017	0,10	473	0,018	0,36	626	0,018	0,59	449	0,017	0,07	1813	0,017	0,10
ʃ	183	0,012	0,20	301	0,011	0,52	398	0,012	0,25	354	0,013	0,03	1236	0,012	0,08
ó	140	0,009	0,65	272	0,010	0,05	307	0,009	0,82	310	0,011	0,15	1029	0,010	0,08
t	639	0,042	0,15	1220	0,046	0,96	1532	0,044	0,025	1207	0,044	0,025	4598	0,044	0,05
č	165	0,011	0,94	343	0,013	0,10	421	0,012	0,04	328	0,012	0,03	1257	0,012	0,025
ɓ	69	0,004	0,02	88	0,003	0,38	92	0,003	0,15	77	0,003	0,73	326	0,003	0,01
k	443	0,029	0,31	614	0,023	0,60	881	0,028	0,40	759	0,028	0,25	2797	0,027	0,007
c	191	0,012	0,10	377	0,014	0,45	425	0,012	0,009	296	0,011	0,27	1287	0,013	0,005
p	447	0,029	0,33	820	0,031	0,01	1051	0,030	0,38	932	0,034	0,55	3250	0,031	0,005
a	1591	0,104	0,86	2481	0,093	0,20	3299	0,095	0,07	2547	0,093	0,20	9918	0,095	0,005
ɨ	114	0,007	0,10	128	0,005	0,31	201	0,006	0,10	143	0,005	0,15	586	0,006	0,005
ń	380	0,025	0,025	775	0,029	0,38	889	0,026	0,12	665	0,024	0,25	2709	0,026	0,004
b	183	0,012	0,07	392	0,015	0,32	486	0,014	0,39	318	0,012	0,06	1379	0,014	0,004
ʒ	36	0,002	0,15	68	0,002	0,002	58	0,002	0,04	44	0,002	0,05	206	0,002	0,002
ź	24	0,002	0,56	29	0,001	0,025	45	0,001	0,13	41	0,002	0,20	139	0,001	0,001
l	311	0,020	0,52	598	0,022	0,001	729	0,021	0,53	573	0,022	0,001	2211	0,021	0,001

Таблица 4 окончание

Фонемы	Я. Ивашкевич «Девушка и голуби» (5 выборок) Σ = 15 319			Л. Кручковский «Первый день свободы» (7 выборок) Σ = 25 689			С. Мрожек Пять рассказов (5 выборок) Σ = 34 606			Е. Шанявский Пять рассказов (5 выборок) Σ = 27 314			Сумма четырех текстов (22 выборки) Σ = 103 828		
	f	p	q_i	f	p	q_i	f	p	q_i	f	p	q_i	f	p	q_i
g	244	0,016	0,11	336	0,013	0,15	505	0,015	0,025	353	0,013	0,08	1438	0,014	0,001
n	606	0,039	0,30	960	0,036	0,0025	1511	0,044	0,57	1058	0,039	0,08	4135	0,040	0,0005
ž	197	0,013	0,32	406	0,015	0,35	471	0,014	0,0005	442	0,016	0,07	1516	0,015	0,0005
x	166	0,011	0,025	310	0,012	0,01	357	0,010	0,08	266	0,010	0,03	1099	0,010	0
ṕ	44	0,003	0,02	96	0,004	0,01	97	0,003	0,005	91	0,003	0	328	0,003	0
f́	17	0,001	1	42	0,001	0,01	67	0,002	0,003	58	0,002	1	126	0,001	0
v	372	0,024	0	588	0,022	0,40	952	0,027	0,11	693	0,025	0,15	2605	0,025	0
š	280	0,018	0,22	638	0,024	0,002	764	0,022	0,10	518	0,019	0,02	2200	0,021	0
ś	293	0,019	0,15	559	0,021	0,18	646	0,019	0,28	443	0,016	0,025	1941	0,019	0
ź	119	0,008	0,08	263	0,010	0,89	196	0,006	0,002	219	0,008	0,22	797	0,008	0
ć	169	0,011	0,52	525	0,020	0,60	551	0,016	0,80	355	0,013	0,09	1600	0,015	0
d	392	0,026	0,95	519	0,020	0,05	858	0,025	0,52	627	0,023	0	2396	0,023	0
µ	511	0,033	0,04	408	0,015	0,003	1063	0,031	0,50	778	0,028	0,025	2760	0,027	0
j	385	0,025	0,04	838	0,031	0,05	775	0,022	0,01	696	0,025	0,004	2694	0,026	0
r	478	0,031	0,40	737	0,028	0,30	1087	0,032	0,005	871	0,032	0,002	3173	0,030	0
m	418	0,027	0,45	997	0,037	0,32	1128	0,033	0	988	0,036	0,33	3581	0,034	0
õ	74	0,005	0,16	177	0,006	0,38	195	0,006	0,008	186	0,007	0	632	0,006	0
i	1327	0,087	0,49	2023	0,076	0,60	2802	0,081	0,38	2259	0,083	0,35	8411	0,081	0
e	1569	0,102	0,52	3157	0,119	0,77	3555	0,103	0,01	2992	0,110	0,03	11273	0,109	0

Примечание: f — абсолютная частота; p — относительная частота; q_i — вероятность, с которой табличное χ^2_q превысит экспериментальное χ^2.

Таблица 5

Группировка фонем соответственно интервалам значений, полученным при проверке по критерию χ^2 для альтернативного признака

Интервалы	Я. Ивашкевич	Л. Кручковский	С. Мрожек	Е. Шанявский
1—0,60	*sēóčadǰ*	*sģńntkź ćie*	*uóć*	*scb̨nǰ*
0,59—0,21	*ģńoukp źlnžš ćrmie*	*ēouzfb̨ ckńbžv rmō*	*sozfk pblnś dụi*	*kopń źmi*
0,20—0,11	*ftzgšō*	*agś*	*ēb̨ńźv*	*ńv̆akźv*
0,10—0,06	*bźzck*	*č*	*axģkš*	*zbgžć*
0,05—0,01	*bńxpụj*	*ópźxpǰ dj*	*čtzóǰ jeń*	*ģuftčx sśụezē*
0,009—0,0005	—	*źlnšụ*	*cžpzrō*	*ljr*
0	*vž*	*ž*	*mž*	*dpžō*

с соответствующими фонемами в других текстах. Для удобства сравнения расположим данные, приведенные в табл. 4, немного иначе и сгруппируем их в табл. 5.

Критическая область критерия устанавливается начиная с $q = 0{,}05$. Тогда получаем, что в первом тексте в критическую область попадают частоты 8 фонем, во втором — частоты 14 фонем, в третьем — частоты 16 фонем и в четвертом — частоты 19 фонем. Таким образом, в каждом тексте частоты большей части фонем оказываются достаточно стабильными, чтобы не отвергалась гипотеза об однородности. Частоты трех фонем *sok* в каждом тексте оказываются настолько стабильными, что для них значение q_i попадает в область, ограниченную снизу 0,21, т. е. является весьма высоким. Частоты следующих фонем в каждом тексте попадают в область допустимых значений ($q > 0{,}05$): *szýbkǩćaio*.

Рассмотрим частоты фонем во всем исследованном тексте ($N = 103\,828$ фонем). При проверке этого текста на однородность по отношению к частотам фонем по критерию χ^2 для альтернативного признака устанавливаем более низкий уровень значимости $q = 0{,}0005$, соответствующий увеличению объема выборки. Получаем, что в область допустимых значений $q > 0{,}0005$ попадают следующие 23 фонемы: *sē ģńouzfótčbkcpakńbźzlg*. И здесь большая часть фонем имеет достаточно стабильную частоту, чтобы не отвергнуть гипотезу об однородности. Сравним оба списка фонем и выделим в них общую часть: *szýbkǩao*. Можно считать, что в рамках нашего эксперимента и в терминах применяемого критерия частота каждой из этих восьми фонем не противоречит гипотезе об однородности. Это утверждение может быть сделано,

поскольку мы сопоставили данные по каждому из текстов с данными по объединенному тексту и выбрали только совпадающие результаты. Это дает возможность осуществить своего рода перекрестную проверку результатов. Кроме того, напомним, что критерий χ^2 является весьма сильным, и неотвержение гипотезы об однородности при проверке частот всего одной фонемы свидетельствует о действительной стабильности этой частоты.

Разумеется, получившиеся результаты следует интерпретировать статистически, т. е. иметь в виду возможность того, что в другом эксперименте подобную стабильность продемонстрируют другие фонемы. Особенно это касается фонемы *ќ*. По-видимому, ее появление в этом списке объясняется случайностью. Что же касается фонем *szao kbó* (а особенно *szoka*, имея в виду наибольшую последовательность результатов именно по этим фонемам), то можно с определенной достаточно большой вероятностью предположить, что в и других текстах их частота будет стабильной.

Среди этих фонем мы находим гласные низкой (*a*, *o*), но не высокой тональности. Смычные согласные относятся к периферийным согласным также низкой тональности (*kb*), а непрерывные *sz*, напротив, относятся к непериферийным согласным и составляют минимальную бинарную фонологическую оппозицию.

Рассмотрим теперь фонемы, частота которых противоречит гипотезе об однородности. Здесь труднее выделить сколько-нибудь постоянную группу таких фонем, чем в случае стабильности частоты. Всего обнаруживается четыре фонемы — *ʒxjṕ*, которые в каждом тексте имеют нестабильную частоту. Эти же фонемы имеют $q_i = 0$ при проверке их частоты по критерию χ^2 для альтернативного признака в объединенном тексте. Приблизительно можно выделить группу фонем *rlụmx*, частота которых имеет тенденцию быть нестабильной. Правда, *rlmụ* встречаются и в области допустимых значений, однако чаще (примерно в 3 случаях из 4) частота этих фонем нестабильна; в объединенном тексте *rmụ* имеют $q_i = 0$. Таким образом, в группу фонем с нестабильной частотой попадают фонемы, идентифицируемые как «негласные» и «несогласные», а также *x*, не имеющий звонкого коррелята. Что же касается фонем *ʒ* и *ṕ*, то они попали в группу фонем, имеющих тенденцию к нестабильной частоте по той же причине, по которой *ẽ* и *ġ* оказались в объединенном тексте в группе фонем с весьма стабильной частотой: и те и другие имеют слишком малую частоту, чтобы результаты можно было считать окончательными и достоверными.

О чем говорят полученные результаты применения критерия проверки χ^2 для установления однородности текстов относительно частот отдельных фонем?

Прежде всего о том, что, исследуя однородность текстов относительно частот отдельных фонем, нельзя составить однозначную картину статистической структуры текста на фонологическом уров-

не. Ни один из результатов не является решающим: две выборки из текста могут быть равным образом однородны и неоднородны относительно частот всего ряда фонем. С другой стороны, тексты оказываются однородными в отношении частот одних фонем и неоднородными в отношении частот других, причем каждый раз по-своему: весь инвентарь фонем, использованных для порождения одного данного текста, может быть разделен на две части — фонемы, имеющие стабильную частоту в смысле критерия χ^2, и фонемы, имеющие нестабильную частоту. Можно утверждать, что любой текст будет построен подобным образом. Несмотря на то, что удается установить определенные зависимости в поведении текста по отношению к частотному распределению фонем, связанные с объемом и законченностью текста, а также выделить приблизительные группы фонем, имеющие тенденцию к стабильной или нестабильной частоте, эти зависимости и характеристики оказываются весьма грубыми и ориентированными не на фонологическую систему, а на текст.

Иными словами, оказывается, что фонемы образуют такую же статистическую совокупность, что и словарь: их поведение значимым образом зависит от характеристик текста; так же, как и слова, фонемы употребляются не только в силу чистой случайности (как полагал Г. Хердан), но и в результате направленного выбора. Именно поэтому в любом тексте обязательно будут фонемы, частота которых нестабильна — это соответствует тому, что фонема представляет собою не мельчайшую неразложимую единицу языка (как утверждает Г. Хердан), но сложную структуру, поведение которой управляется различными факторами, в том числе и выбором. Если бы все фонемы имели совершенно стабильную частоту, гибкость языкового кода была бы ограничена, поскольку она определяется не априорной частотой элемента, но возможностью существенных отклонений от этой частоты.

Таким образом, в неоднородности текстов относительно частот фонем (причем каждый текст неоднороден относительно частот фонем, образующих, по-видимому, нигде не повторяющийся класс, ср. табл. 4) мы видим явление, изоморфное подобной же неоднородности на словарном уровне. Резкое противопоставление фонологического и лексического уровней, которое постулировал в статистическом плане Г. Хердан, оказывается ошибочным. Можно ли на этом поставить точку и заявить, что, поскольку фонемы на уровне статистики ведут себя подобно словам, демонстрируя нестабильную частоту, проблема статистической структуры фонологического уровня является снятой?

Подобное решение было бы справедливым, если бы все фонемы во всех экспериментах регулярно имели нестабильную частоту. Между тем, как мы только что видели, в каждом тексте больше половины фонем имеет стабильную частоту в терминах весьма сильного критерия, при этом в распределении фонем по стабиль-

ным и нестабильным группам можно усмотреть пусть приблизительную закономерность.

Дело в том, что статистическая структура оказывается, по-видимому, более сложной, чем предполагалось раньше. Распределение фонем по стабильным и нестабильным в смысле частоты группам отдаленно отражает эту глубокую структуру, вскрытие которой является одновременно и более трудной, и более интересной задачей, чем это допускали лингвисты, применявшие статистические методы. Собственно говоря, дело обстоит подобным образом и в отношении статистической структуры словаря: если прежде закон Ципфа (в разных модификациях) формулировался для всего словника, то теперь стало очевидным, что словарь необходимо дифференцировать, и что закон Ципфа в его традиционной формулировке справедлив только для слов, лежащих в интервале $50 < r < 1500$ [39].

Если не удается найти четкой закономерности в распределении фонем в текстах группы стабильной (или нестабильной) частоты, следует изменить сам подход к нахождению статистической структуры. Эту структуру следует искать не на уровне фонем, а на более глубоком уровне различительных признаков. Соответственно мы принимаем гипотезу, что текст является совокупностью подобных дискретных признаков, а также что статистическое поведение фонем отражает закономерности статистической структуры, образуемой различительными признаками, модифицированные под влиянием тех специфических факторов, которые несет каждый отдельный текст.

§ 3. Эксперименты по проверке текстов на однородность относительно частот классов фонем

1. Поскольку на предыдущем этапе эксперимента было установлено, что поведение фонем в смысле стабильности их частоты в текстах зависит от самого текста, нашей задачей на настоящем этапе было нахождение таких единиц фонологического уровня, частота которых отличалась бы стабильностью вне зависимости от текста. Естественно, что первым шагом явилась проверка однородности текста относительно частот элементарных различительных признаков, из которых строится фонологическая система польского языка (см. табл. 1). Подобный эксперимент был проделан с использованием критерия χ^2 для альтернативного признака на тексте Я. Ивашкевича. Частота признака определялась по суммарной частоте всех фонем, идентифицируемым этим признаком, например $\text{p}(\text{«гласность»}) = \text{p}(a) + \text{p}(o) + \text{p}(e) + \text{p}(i) + \text{p}(u) + \text{p}(\tilde{o}) + \text{p}(\tilde{e})$.

[39] Ср. цитировавшиеся выше работы Р. М. Фрумкиной и Д. М. Сегала на эту тему.

В результате проверки мы получили довольно высокие значения q_i для некоторых признаков:

q_i (гласность)	=0,975	q_i (периферийность)	=0,80
q_i (согласность)	=0,95	q_i (яркость)	=0,70
q_i (компактность)	=0,30		

Однако для остальных признаков эти значения были довольно низкими:

q_i (диффузность)	=0,04	q_i (звонкость)	=0,10
q_i (непрерывность)	=0,015	q_i (палатальность)	=0,005
q_i (назальность)	=0,02		

Хотя для этого случая мы установили уровень значимости, равный 0, 05, однако, поскольку объекты нашего подсчета имеют относительную частоту гораздо большего порядка, чем фонемы, желательно получить высокое значение q_i для решительного неотвержения исходной гипотезы,— поэтому q_i для звонкости, равное 0, 10, все же показывает довольно слабое согласие.

Таким образом, уже после предварительного эксперимента выяснилось, что подобный механический подход к определению элементов фонологического уровня со стабильной частотой неправилен. Дело в том, что как по своей дистрибуции, так и по объему различительные признаки неравнозначны. Одни из них членят всю совокупность действительно на две примерно равные доли (гласность—негласность), так что применение критерия для альтернативного признака оправдано — каждая фонема либо гласная, либо негласная. Для других признаков это явно не так: например, не каждая фонема получает плюс или минус по признакам «непрерывность», «звонкость», «палатальность», «назальность», «диффузность». Эти признаки, следовательно, нельзя считать строго альтернативными. Другие признаки идентифицируют лишь небольшое число фонем и проч.

Таким образом, для того, чтобы иметь возможность вскрыть статистическую структуру на фонологическом уровне, необходимо найти разбиение рассматриваемой совокупности на некоторые непересекающиеся классы. При этом каждый класс будет репрезентировать определенную единицу фонологического уровня, меньшую чем фонема (в том смысле, что фонема образуется на пересечении нескольких подобных единиц), и в свою очередь образующуюся на пересечении нескольких различительных признаков.

Естественно, что множество из 42 фонем можно было разбить на огромное количество подмножеств ($n = 2^{42}$), и следовало выбрать наиболее фонологически осмысленные и статистически устойчивые классы.

2. Мы избрали следующую эвристическую процедуру, близкую процедурам линейного программирования: сначала все множество фонем разбивается на два больших класса по наиболее общему различительному признаку «гласность—негласность», частоты этих классов близки одна другой и при предварительной проверке весьма устойчивы. Затем проводится следующее по порядку дихотомическое деление — класс негласных разбивается на класс согласных и класс несогласных (т. е. сонантов и плавных). Это разбиение производится по признаку «согласность—несогласность», следующему по порядку в таблице после первого признака «гласность». Класс гласных более не дробится, поскольку, как показала предварительная проверка по критерию χ^2 для альтернативного признака на текстах Я. Ивашкевича и Л. Кручковского, членение гласных на классы по любому признаку дает весьма нестабильные частоты. Согласные фонемы в следующем дихотомическом членении делятся на периферийные и непериферийные, что дает два класса с соизмеримой частотой. Разбиение согласных именно по признаку «периферийность», а не по признаку «компактность» было выбрано вследствие того, что предварительная проверка на однородность по критерию χ^2 для альтернативного признака на текстах Я. Ивашкевича и Л. Кручковского показала, что классы, объединяемые признаком «периферийность» (т. е. периферийные фонемы и периферийные согласные фонемы), дают гораздо более стабильные частоты, чем классы, выделяемые по признаку «компактность» (т. е. компактные фонемы и компактные согласные фонемы).

Введение предварительной проверки по критерию χ^2 для альтернативного признака было необходимо, так как требовалось ввести определенную иерархию в порядок признаков, по которым производилось последовательное дихотомическое членение совокупности фонем. Одним из критериев такой иерархии было естественно избрать получение при членении классов с более стабильной частотой. Именно поэтому такие признаки как «палатальность» и «звонкость» никак не могли быть избраны в качестве первых критериев членения совокупности. Другим критерием мы избрали близость частот классов, получающихся после каждого деления. В случае признаков «диффузность» или «яркость» частоты «диффузных» или «ярких» значительно ниже, чем частоты «недиффузных» или «неярких» (0,114—0,204 и 0,050—0,123), поэтому их нельзя было избрать для первоначального членения. Помимо этого для первоначального членения следует избрать признаки, по которым фонемы не получают (или почти не получают) нулей в матрице идентификации. Вследствие всех этих причин порядок первых членений был таков, как описано выше: гласность, согласность, периферийность.

При следующем членении в качестве основного критерия мы опять избрали получение наибольшего значения q_i (т. е. наиболь-

шей стабильности частот при предварительной проверке на двух текстах по критерию χ^2 для альтернативного признака). В результате этой проверки выяснилось, что если на следующем шаге к обоим классам согласных фонем, получившимся в результате предыдущего членения (периферийным и непериферийным), в качестве критерия членения приложить одинаковый признак (т. е. либо компактность, что дает классы *pṕbb́ff vv́*; *kḱgǵx*; *šš́ž́čč́ǯ́ǯ́*; *tdczsz*, либо непрерывность, что дает классы *pṕbb́kḱgǵ*; *ff vv́x*; *sśšźž́*; *tdczć́ǯ́č́ǯ́*), то следующие классы фонем регулярно имеют весьма нестабильную частоту: *šśž́č́ć́ǯ́ǯ́*; *tdczsz*; *ff vv́x*; *pṕbb́kḱgǵ*, в то время как классы *pṕbb́ff vv́*; *kḱgǵx*; *sśšźž́*; *tdczć́ǯ́č́ǯ́* дают стабильную частоту. Из этого был сделан вывод, что после членения по признаку «периферийность» следующее членение следует производить по разным признакам для обоих классов — для периферийных — это «компактность», а для непериферийных — это «непрерывность». Подобное членение кроме статистических обоснований имеет и чисто фонологический резон.

Здесь уместно вспомнить о понятии естественного класса (natural class), о котором рассказывал в одном из своих докладов в Институте славяноведения в Москве Манфред Бирвиш. Под естественным классом понимается такой класс, для идентификации которого требуется меньше признаков, чем для идентификации каждого из его членов, например для идентификации класса *ui* в польской фонологической системе требуется всего один признак — «диффузность», в то время как каждый член этого класса идентифицируется четырьмя признаками. С другой стороны, для идентификации класса *ǯ̄ōx* требуется 12 признаков — гласность, некомпактность, недиффузность, периферийность, назальность, согласность, непериферийность, компактность, яркость, звонкость, палатальность, непрерывность — явно больше, чем для любого из членов этого класса.

Нам представляется, что членение на классы *pṕbb́ff vv́*; *kḱgǵx*; *sśšźž́*; *tdczć́ǯ́č́* является в этом смысле более естественным, чем любое другое. Дело в том, что членение всех согласных на периферийные и непериферийные, будучи фундаментальным в качестве первичного этапа, объединяет вместе согласные противоположных мест образования: губные и заднеязычные. Естественно, что задачей следующего по порядку членения является разделение этих двух классов (которые могут быть отождествлены лишь указанием своего места образования, следовательно, всего одним признаком).

Непериферийные согласные обладают бо́льшим единством в смысле места образования, чем периферийные, — условно к ним можно приложить термин «язычные» согласные. Следовательно, эти согласные можно членить по признаку способа образования. Кроме того, членение непериферийных согласных именно по способу образования дает близкие частоты (0,102 и 0,117 для всего текста).

И, наконец, среди периферийных согласных признак непрерывности релевантен главным образом для некомпактных (губных), поэтому сначала требовалось выделить более равномерные классы. В то же время среди периферийных согласных непрерывные образуют большой класс, куда равномерно входят как компактные, так и некомпактные члены.

Следующее разбиение касается несогласных наряду с согласными. Несогласные разбиваются по признаку «носовость—ртовость», дающему наиболее естественные классы несогласных; сонанты *mḿ nń* и плавные *rlu̯j*, каждый из которых может быть идентифицирован всего лишь одним признаком и частота которых близка (0,108 и 0,104 по всему тексту).

Членение согласных снова осуществляется по нескольким параллельным признакам: периферийные компактные и некомпактные делятся на прерывные и непрерывные, давая классы *pṕbb́*; *ffv́*; *kḱgǵ*, *x*. Периферийные делятся неравномерно: класс *sśšzźž* не членится, а класс *tdszć́ź̌č̌ʒ́* членится по признаку «яркость — тусклость» на классы *td* и *cć̆čʒ́ʒ́̌*.

Именно такое членение: *гласные*, *mḿnń*, *rlu̯j*, *pṕbb́*, *ffv́*, *kḱgǵ*, *x*, *sśšzźž*, *td* и *ćc̆ć́ʒ́ʒ́̌* продемонстрировало наибольшую стабильность частот при предварительной проверке по критерию χ^2 для альтернативного признака на тексте Я. Ивашкевича. Разумеется, если текст делить на меньшее количество более крупных классов, то стабильность будет большей. Данное членение является максимальным возможным при сохранении однородности; если дробить имеющиеся классы на более дробные (по звонкости, палатальности, компактности), то картина резко меняется и однородность исчезает. Любое другое фонологически значимое членение давало классы с гораздо менее стабильной частотой. Отметим, что и при данном членении частота класса *ffv́* при разбиении текста Я. Ивашкевича на наши пять зон противоречит гипотезе об однородности ($q_i = 0{,}001$ при выбранном для предварительной проверки уровне значимости $q = 0{,}05$). Однако, если брать другое максимальное допустимое при стремлении сохранить однородность членение фонем на классы, количество неоднородных классов будет большим.

Приведем некоторые данные в этом отношении: q_i для класса *szffv́* при предварительной проверке равно 0, для класса *ffv́x* q_i также равно 0.

Было специально составлено несколько классов так, чтобы входящие в них члены имели как можно меньше общих признаков, т. е. образовывались «неестественные классы», и их частота проверялась на однородность по критерию χ^2 для альтернативного признака в тексте Я. Ивашкевича.

Результаты вполне соответствуют ожиданиям: q_i для класса (*rldtspšžkž*) равно 0; q_i для класса (*mnju̯tdczpb̆vśšgxõia*) равно 0;

q_i для класса (*xõšlkd*) равно 0; q_i для класса (*fvzɯkš*) равно 0; q_i для класса (*źdfxkpṕcǵ*) равно 0; q_i для класса (*šfkpdsvv́*) равно также 0.

Разумеется, мы не могли перебрать все подобные классы; возможно, что для какого-нибудь из них q_i будет лежать и в области допустимых значений, однако можно с определенной уверенностью утверждать, что полученное нами членение множества фонем на классы является значимым и в некотором смысле единственным, так как оно дает наивысшие шансы на однородность. С другой стороны, выявляется тот факт, что случайные объединения фонем, не являющиеся естественными классами, имеют тенденцию к крайне нестабильной частоте. Эта тенденция справедлива не только для классов, куда входят фонемы, имеющие минимальное число общих признаков, но и для классов, куда входят фонемы с противоположными значениями одного признака, т. е. q_i для класса, объединяющего периферийные компактные с непериферийными компактными (*pṕbb́ffv́*, *šśžźčćžź́*), равно 0,002; q_i для класса, объединяющего непериферийные некомпактные с периферийными компактными (*tdczszkǩgǵx*), равно 0; q_i для класса, объединяющего непериферийные некомпактные прерывные с периферийными компактными прерывными (*tdczpṕbb́*), равно 0 025; q_i для класса, объединяющего непериферийные некомпактные прерывные с периферийным компактным непрерывным *x* (*tdczx*), равно 0; q_i для класса, объединяющего непериферийные компактные непрерывные с периферийными некомпактными прерывными (*ššźžpṕbb́*), равно 0,001.

Все изложенное выше о процедуре членения фонем на классы не является дефинитивным экспериментом по проверке однородности частот классов фонем (описание этого эксперимента последует ниже). Приведенные данные были необходимы в первую очередь для получения наиболее отвечающего условиям эксперимента членения фонем на классы. Получившиеся классы вполне соответствуют некоторому интуитивному представлению о том, как членится система польских фонем. Разумеется, мы могли бы априори избрать подобное членение, никак его не мотивируя статистически. В этом случае, однако, возникал бы вопрос о том, почему избрано данное членение, поскольку имеющаяся система различительных признаков позволяет членить систему фонем другими способами. Тот факт, что в предварительном эксперименте интуитивно наиболее адекватное членение совпадает с членением, дающим классы с наиболее стабильной частотой, представляется заслуживающим внимания. Очевидно, что естественные фонологические классы ведут себя в статистическом плане иначе, чем случайные объединения фонем. Они образуют определенные целостные объединения, функционирующие во многом как элементарные единицы.

Отметим еще одну особенность членения системы фонем, полученного после предварительного эксперимента. Отношения между получившимися классами неоднородны. Внутри периферийных мы получаем бинарное членение сначала по признаку компактности, а затем непрерывности. Отношения здесь в принципе дихотомичны: *pṕbḃ* относится к *fv́* так же, как *kḱgǵ* относится к *x*. Правда, эта симметрия нарушена тем, что в то время как для каждого члена из класса *pṕbḃ* имеется свой коррелят в классе *fv́*, всему классу *kḱgǵ* соответствует лишь один член *x*. Это ставит *x* в особое положение в системе, на чем мы подробнее остановимся специально.

Что же касается непериферийных, то здесь можно было бы, наверное, ввести бинарные отношения, однако в этом случае мы не получаем классов со стабильной частотой. В полученном членении выявляется принципиально троичная система отношений:

sśšzźž
td
cćčzźž

С другой стороны, троичное отношение между классами соответствует троичным отношениям между членами классов:

s	*ś*	*z*	*ź*	*c*	*ć*	*ʒ*	*ź*
	š		*ž*		*č*		*ǯ*

Выделение троичных отношений в подсистеме польских непериферийных согласных соответствует интуитивному и традиционному представлению. Троичное противопоставление классов *td* — *sśšzźž* — *cćčzźǯ* может быть описано как противопоставление по прерывности—непрерывности: класс *cćčzźǯ* будет идентифицирован как обладающий и не обладающий данным признаком[40], что вполне соотносится с традиционной характеристикой аффрикат как фонем, обладающих одновременно свойствами взрывных и фрикативных.

Противопоставление внутри классов может быть описано либо как противопоставление по палатализованности—непалатализованности (шипящие *š*, *ž* и аффрикаты *č*, *ǯ* характеризуются как обладающие и не обладающие этим признаком, что соответствует фактам исторической морфонологии польского языка, ср. mucha — musze), либо как противопоставление по компактности—диффузности, при этом *śźćź* идентифицируются как одновременно компактные и диффузные.

Тот факт, что в пределах одной фонологической системы соседствуют в одной подсистеме преимущественно двоичные, а в

[40] См.: М. И. Лекомцева. К описанию фонологической системы старославянского языка на основе тернарного принципа. «Лингвистические исследования по общей и славянской типологии». М., 1966, стр. 122.

другой — троичные отношения, показывает внутреннюю динамичность системы, возможность изменений. Этот факт указывает также на отсутствие строгой симметрии внутри фонологических систем (ср. во Введении о проявляющемся в последние годы внимании к периферийным участкам системы).

Отметим, что в теоретическом плане вопрос о возможном сосуществовании внутри одной системы подсистем, основывающихся на различных отношениях, близок к идеям о возможности n-арных описаний, высказанным в только что упоминавшейся нами (в сноске) работе М. И. Лекомцевой.

* * *

1. Определив оптимальное разбиение фонологической системы на классы, мы предприняли центральный эксперимент настоящей работы — установление однородности текстов по отношению к частотному распределению классов фонем. Проводя подобный эксперимент с фонемами, мы были ограничены рассмотрением лишь двух выборок за один раз. Это связано, с одной стороны, с тем, что вычисления, применяемые в критерии χ^2 для сравнения между собою n выборок по m признакам (а не по одному, как в критерии для альтернативного признака), как это станет ясно из нижеследующего, слишком трудоемки при наличии 42 разрядов сравнения. С другой стороны, дефинитивные результаты были получены и при попарном сравнении выборок.

В случае частотного распределения классов фонем количество членов сравнения не слишком велико, и при помощи настольного электронного калькулятора оказалось возможным провести соответствующие вычисления в обозримое время.

Критерий χ^2 для сравнения нескольких вероятностей в нескольких группах опытов (будем называть его «многомерным» критерием χ^2) применяется в следующей ситуации: пусть некоторое событие в первых n_1 опытах наступило x_1 раз и не наступило y_1 раз, во второй группе из n_2 опытов оно наступило x_2 раз и не наступило y_2 раз, и т. д. Нужно проверить, одинаковы ли вероятности этого события во всех группах опытов.

Еще более общим является следующий случай, который и описывает конкретную ситуацию, возникающую в наших экспериментах. Пусть n_1 объектов (в нашем случае — количество фонем в первой выборке) разбиты по какому-то признаку на h классов (в нашем случае — количество классов фонем) и пусть $x_1, y_1, \ldots$ $\ldots z_1$ — количество объектов в этих классах (абсолютная частота каждого класса). Следующим n_2 объектам (объем второй выборки) аналогичным образом соответствуют количества $x_2, y_2, \ldots z_2$ и т. д. до $x_k, y_k, \ldots z_k$.

Таким образом, в результате наблюдений получается hk чисел (где k, в нашем случае,— общее количество выборок), которые

располагаются в прямоугольную схему:

h классов	p	x_1	x_2	. . .	x_k	Σx
	q	y_1	y_2	. . .	y_k	Σy
	.	.	.		.	.
	.	.	.		.	.
	.	.	.		.	.
	r	z_1	z_2	. . .	z_k	Σz
		n_1	n_2	. . .	n_k	N

Нужно проверить, могут ли вероятности p, q, . . . r, соответствующие h классам, быть одинаковыми для каждого из k столбцов.

Обозначим через i номер ряда (класса фонем), j — номер выборки; через V_{ij} — абсолютную частоту i-ого класса фонем в j-ой выборке, через V_j — сумму частот класса фонем в i-ом ряду (x, y, . . ., z), через V_j — объем j-ой выборки. N — общий объем текста.

Формула многомерного критерия χ^2 выглядит следующим образом:

$$\chi^2 = N \cdot \left(\sum_{ij} \frac{V_{ij}^2}{V_i V_j} - 1 \right).$$

Таким образом, основная трудность в применении этого критерия состоит в суммировании по большому количеству разрядов.

Число степеней свободы равно $(h-1)(k-1)$, т. е. произведению числа классов фонем минус 1 на число выборок минус 1.

2. Эксперимент состоял из нескольких стадий в соответствии с последовательными членениями всего инвентаря фонем на классы.

Первая стадия

На этой стадии проверялась однородность каждого из четырех текстов, а также всего текста в целом относительно распределения частот гласных и негласных фонем. Здесь все множество фонем членится лишь на два самых общих класса. Данные по этому эксперименту см. в табл. 6—9.

Таким образом, мы убеждаемся в том, что части одного текста, а также различные тексты одного автора оказываются не противоречащими гипотезе об однородности частот гласных и негласных фонем при уровне значимости $q = 0{,}05$.

В данном эксперименте был избран именно этот уровень значимости, поскольку относительные частоты сами по себе не являются слишком малыми, а выборки не слишком велики, и поскольку сравнивались лишь два класса.

Таблица 6

Текст Я. Ивашкевича

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	1329	1310	1319	1322	1158	6 438
Негласные	1819	1839	1786	1829	1608	8 881
V_j	3148	3149	3105	3151	2766	N=15 319

$\chi^2 = 0{,}5751$; $k = (2-1)(5-1) = 4$; $q_{i\,(k=4)}(\chi_q^2 > \chi^2) = 0{,}95$

Таблица 7

Текст Л. Кручковского

Класс	Выборка							
	I	II	III	IV	V	VI	VII	V_i
Гласные	1399	1437	1518	1506	1469	1502	2082	10 913
Негласные	2020	2147	2200	2150	2098	2157	2904	15 676
V_j	3419	3584	3718	3656	3567	3659	4986	N = =26 589

$\chi^2 = 12{,}1378$; $k = (2-1)(7-1) = 6$; $q_{i\,(k=6)}(\chi_q^2 > \chi^2) = 0{,}05$

Таблица 8

Текст Е. Шанявского

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	1577	2196	2332	2555	2647	11 307
Негласные	2232	3000	3351	3657	3767	16 007
V_j	3809	5196	5683	6212	6414	N=27314

$\chi^2 = 2{,}1397$; $k = 4$; $q_{i\,(k=4)}(\chi_q^2 > \chi^2) = 0{,}70$

Таблица 9

Текст С. Мрожека

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	2018	2629	2818	3057	3563	14 085
Негласные	2856	3798	4113	4502	5252	20 521
V_j	4874	6427	6931	7559	8815	N=34 606

$\chi^2 = 2{,}0632$; $k = 4$; $q_{i\,(k=4)}(\chi_q^2 > \chi^2) = 0{,}70$

Таблица 10

Объединенный текст

Класс	Текст				
	Я. Ивашкевич	Л. Кручковский	Е. Шанявский	С. Мрожек	V_i
Гласные	6 438	10 913	11 307	14 085	42 743
Согласные	8 881	15 676	16 007	20 521	61 085
V_j	15 319	26 589	27 314	34 606	103 828

$\chi^2 = 8{,}52$; $k = (4—1)(2—1) = 3$; $q_{i\,(k=4)}(\chi_q^2 > \chi^2) = 0{,}035$.

Соответственно каждый из четырех текстов можно рассматривать как одну однородную внутри себя совокупность. Сравним наши четыре текста (табл. 10).

Как видим, здесь значение q_i понижается до 0,035. Это вполне естественно, поскольку мы имеем дело со случайной совокупностью законченных текстов.

Однако нам представляется закономерным в этом случае понизить уровень значимости, чтобы учесть тот факт, что теперь мы имеем дело с гораздо более крупными выборками, чем в предыдущем эксперименте. По-видимому, уровень значимости $q = 0{,}01$ будет, с одной стороны, достаточно высоким для двух разрядов, а с другой стороны, позволит сделать поправку на весьма большую величину N (именно эта величина прямым образом влияет на результат при увеличении выборки).

Таким образом, мы можем считать первым позитивным результатом нашего исследования неотвержение гипотезы об однородности замкнутого польского текста, относимого к художественной литературе, относительно распределения частот гласных и негласных фонем.

Иными словами, статистически установлено, что в законченном художественном тексте на польском языке доля гласных и негласных всегда будет одинакова, а отклонения не выйдут за пределы случайных.

Настоящий результат имеет значение не только для выявления статистически стабильных элементов в фонологической системе языка, но и для характеристики структуры связного текста. Постоянство содержания гласных и негласных в текстах, которые заведомо не образованы как статистические случайные выборки, является довольно нетривиальным фактом.

Таким образом, обнаружен первый элемент статистической структуры на фонологическом уровне: стабильность фонологических частот существует, но это не стабильность частот фонем, а стабильность частот единиц более глубокого уровня.

Здесь следует отметить, что для украинского языка однородность текстов относительно частот гласных и негласных фонем внутри текстов, относящихся к одному функциональному стилю, была доказана в недавней работе Л. М. Гридневой [41].

Эта работа — единственная из известных автору, в которой анализируется проблема статистической однородности на фонологическом уровне. Выполненная на большом материале (выборки объемом по 50 000 фонем из шести функциональных стилей: драматургия, художественная проза, поэзия, устная речь, публицистика и научная проза) с привлечением электронно-счетной техники, эта работа касается лишь одного аспекта статистической однородности — однородности относительно частот гласных, негласных и пропусков. Но в этом аспекте исследование проводилось на самом высоком уровне, с привлечением всего необходимого статистического аппарата (применялся тот же многомерный критерий χ^2, что и в настоящей работе), поэтому результаты можно считать абсолютно достоверными и дефинитивными. Подсчитаны значения q_i для каждого стиля (текст каждого стиля разделялся на 10 выборок каждая объемом в 5000 фонем). Эти значения равны: для драматургии 0,2560; для художественной прозы 0,3142; для поэзии 0,9694; для устной речи 0,9694; для публицистики 0,5745; для научной прозы 0,9813. Все эти значения не противоречат гипотезе о статистической однородности текста относительно распределения частоты гласных и негласных фонем.

Л. М. Гриднева проводит также проверку предположения о нормальности распределения частот гласных, негласных и пропусков в тексте. Строится график эмпирического распределения, форма которого оказывается близкой к форме теоретической кривой. По формуле кривой нормального распределения вычис-

---

[41] Л. М. Гриднева. Розподіл голосних, приголосних, пропусків у сучасному українському мовленні. «Статистичні та структурні лінгвістичні моделі». Київ, 1966, стр. 45—53.

ляются теоретические частоты, и степень близости между эмпирическими и теоретическими частотами проверяется при помощи критерия χ^2. Вычисленные значения χ^2 и соответствующие им значения q_i показывают, что гласные, негласные и пропуски распределены по нормальному закону. Можно, следовательно, утверждать, что наиболее вероятная частота гласных, негласных и пропусков равна средней частоте этих классов в соответствующих стилях.

Таким образом, мы видим, что вывод об однородности текста относительно распределения частоты гласных и негласных подкрепляется обнаружением нормальности этого распределения.

К сожалению, в работе Л. М. Гридневой не проверяется однородность всей генеральной выборки из шести функциональных стилей. Насколько можно судить, автор склоняется к точке зрения, что каждый стиль представляет собой отдельную генеральную совокупность, отграниченную от других стилей: «Можно установить, существенны ли расхождения между средними в разных стилях, чтобы использовать выявленные закономерности для размежевания разных стилей речи.

Таким образом, статистические вычисления позволяют подтвердить правомерность объединения нескольких однотипных текстов в один стиль» [42].

Рассмотрим результаты Л. М. Гридневой с той точки зрения, насколько они оправдывают разделение материала на шесть замкнутых функциональных стилей. Однако предварительно необходимо представить экспериментальные данные в несколько ином виде, чем это сделано у Л. М. Гридневой. Окончательные результаты подсчетов Л. М. Гридневой [43] представлены в табл. 11.

Таблица 11

Класс	Стиль					
	Драматургия	Художественная проза	Поэзия	Устная речь	Публицистика	Научная проза
Гласные	350±2,22	351±1,66	340±2,22	355±2,22	360±2,22	361±2,77
Согласные	474±3,6	432±3,6	481±2,77	484±3,32	504±3,05	505±2,27
Пропуски	174±4,7	167±4,15	180±4,15	162±3,32	137±3,32	136±3,88

Совершенно очевидно, что эти результаты определенным образом искажают реальное соотношение гласных и согласных в стилях: в публицистике и научной прозе доля гласных и согласных выше, чем в других стилях, однако это связано с повышением

[42] Л. М. Гриднева. Указ. соч., стр. 53.
[43] См. там же, стр. 52.

длины слова (и, следовательно, с понижением доли пропусков). Таким образом, в качестве основного фактора здесь выступает длина слова, а соотношение гласных и согласных оказывается зависимой величиной.

Со своей стороны, данные о средней длине слова представляют большой интерес: как и следовало ожидать, средняя длина слова является наибольшей в публицистике и научной прозе — стилях, ориентирующихся на письменную речь. Наименьшая длина слов — в поэзии. Однако эти данные, к сожалению, нельзя интерпретировать, поскольку Л. М. Гриднева не сообщает, какой принцип был положен в основу выделения слова. Поэтому мы решили исключить данные о содержании пропусков в украинских текстах и получили следующие цифры реального соотношения гласных и негласных между собой:

Таблица 12

Класс	Стиль					
	Драматургия	Художественная проза	Поэзия	Устная речь	Публицистика	Научная проза
Гласные	0,4248	0,4214	0,4141	0,4231	0,4167	0,4168
Негласные	0,5752	0,5786	0,5859	0,5769	0,5833	0,5832

Как мы видим, в таком представлении картина немного меняется: публицистика и научная проза, оставаясь практически идентичными по своим показателям, приближаются к поэзии — повышение доли гласных в указанных двух стилях оказывается мнимым, на деле процент гласных — ниже, чем в драматургии, художественной прозе и устной речи.

Уточненные результаты показывают, что мнение Л. М. Гридневой о возможности выделения шести стилей на основе сравнения встречаемости гласных и согласных является достаточно спорным.

Обращает на себя внимание близость показателей для драматургии, устной речи и художественной прозы, с одной стороны, и для публицистики, научной прозы и поэзии, с другой стороны. Первые три подразделения соответствуют тому, что можно назвать устным, а вторые три — письменным языком. Близость показателей для драматургии, художественной прозы и устной речи служит статистическим обоснованием (правда, на украинском материале) того, что априори постулировалось нами для польского языка в начале настоящей главы. Художественная проза и драматургия действительно являются представительными «портретами» некоторого нейтрального стиля, ориентирующегося на устный язык (по крайней мере, что касается соотношения гласных и согласных в тексте).

С другой стороны, научная проза, публицистика и поэзия сознательно отталкиваются от устного языка, им присуща определенная маркированность. В то время как художественная проза и драматургия (по крайней мере в их классических, реалистических образцах) стремятся как можно точнее имитировать жизнь, скрыть свою искусственность, научная проза (и лишь условно выделяемая как особый стиль публицистика), а также поэзия сознательно выдвигают эту искусственность, условность на первый план. Этим можно объяснить близость статистической структуры гласных и согласных в тексте для этих стилей.

Таким образом, на уровне самого общего различительного признака «гласность—негласность», как показывают материалы по украинскому языку, невозможно различить все шесть функциональных стилей, априорно выделяемые Л. М. Гридневой. Это ясно даже без применения критерия согласия — настолько близки соответствующие цифры. По-видимому, с помощью этого признака статистически различаются лишь совокупности нейтрального стиля, ориентирующегося на устную речь, и маркированного стиля, ориентирующегося на обработанную письменную речь.

Проанализированные нами данные Л. М. Гридневой дают дополнительное основание для рассмотрения единой совокупности устной речи, художественной прозы и драматургии в качестве некоторого нейтрального стиля, репрезентирующего данный язык.

Вторая стадия

На следующем этапе эксперимента проверялась однородность испытуемых текстов по отношению к частотному распределению классов гласные — согласные — несогласные. Здесь мы уже не располагаем контрольными данными, каковыми на первой стадии служили материалы Л. М. Гридневой.

Таким образом, мы видим (см. табл. 13—16), что даже при довольно высоком уровне значимости $q = 0{,}05$ гипотеза об однородности текстов относительно частотного распределения классов гласных, согласных и несогласных не отвергается.

Следовательно, каждый из текстов можно рассматривать как внутренне однородную совокупность и исследовать однородность объединенного текста, составленного из четырех внутренне однородных выборок (табл. 17). Отметим также, что наблюдается явная тенденция к большей величине q_i в связных текстах, чем в объединениях нескольких текстов (ср. данные об однородности относительно частот фонем).

Принимая, как и в предыдущем случае, уровень значимости $q = 0{,}01$, получаем, что экспериментальное значение χ^2 находится в области допустимых значений, значит гипотеза об однородности для объединенного текста, а следовательно и для всей совокупности, не отвергается.

Таблица 13

Текст Я. Ивашкевича

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	1329	1310	1319	1322	1158	6 438
Согласные	1154	1179	1129	1160	1043	5 665
Несогласные	665	660	657	669	565	3 216
V_j	3148	3149	3105	3151	2766	15 319

$\chi^2 = 1,922$; $k = (2-1)(3-1) = 8$; $q_{i\,(k=8)}(\chi_q^2 > \chi^2) = 0,98$

Таблица 14

Текст Л. Кручковского

Класс	Выборка							
	I	II	III	IV	V	VI	VII	V_i
Гласные	1399	1437	1518	1506	1469	1502	2082	10 913
Согласные	1330	1403	1423	1383	1331	1396	1869	10 135
Несогласные	690	744	777	767	767	761	1035	5541
V_j	3419	3584	3718	3656	3567	3659	4986	26 589

$\chi^{2'} = 5,8049$; $k = (7-1)(3-1) = 12$; $q_{i\,(k=12)}(\chi_q^2 > \chi^2) = 0,92$

Таблица 15

Текст Е. Шанявского

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	1577	2196	2332	2555	2647	11 037
Согласные	1380	1911	2116	2402	2364	10 173
Несогласные	852	1089	1235	1255	1403	5 834
V_j	3809	5196	5683	6212	6414	27 314

$\chi^2 = 13,5032$; $k = 8$; $q_{i\,(k=8)}(\chi_q^2 > \chi^2) = 0,09$.

Таблица 16

Текст С. Мрожека

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	2018	2629	2818	3057	3563	14 085
Согласные	1836	2457	2583	2912	3309	13 097
Несогласные	1020	1341	1530	1590	1943	7 424
V_j	4874	6427	6931	7559	8815	34 606

$\chi^2 = 7{,}9492;\ k = 8;\ q_{i\ (k=8)}\ (\chi_q^2 > \chi^2) = 0{,}45$

Таблица 17

Объединенный текст

Класс	Выборка				
	Я. Ивашкевич	Л. Кручковский	Е. Шанявский	С. Мрожек	V_i
Гласные	6 438	10 913	11 307	14 085	42 743
Согласные	5 665	10 135	10 173	13 097	39 070
Несогласные	3 216	5 541	5 834	7 424	22 015
V_j	15 319	26 589	27 314	34 606	103 828

$\chi^2 = 13{,}8475;\ k = (4-1)(3-1) = 6;\ q_{i\ (k=6)}\ (\chi_q^2 > \chi^2) = 0{,}04$

Итак, обнаружен еще один различительный признак, делящий совокупность фонем на такие классы, содержание которых в каждом тексте будет постоянным. Постоянство частоты гласных, согласных и несогласных в любом тексте отражает неизменность силлабической структуры данного языка, поскольку основными конституирующими элементами, определяющими структуру слога в славянских языках, являются именно эти три класса фонем.

Как и в случае дихотомии гласных—негласных мы обнаруживаем здесь стабильность частот, но не на фонемном, а на субфонемном уровне.

Следующим этапом нашего эксперимента является переход к членению внутри класса согласных.

Третья стадия

На данной стадии классы гласных и несогласных остаются неизменными, а класс согласных делится на периферийные и непериферийные согласные.

Таблица 18

Текст Я. Ивашкевича

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	1329	1310	1319	1322	1158	6438
Несогласные	665	660	657	669	565	3216
Периф. согл.	500	514	472	509	444	2439
Непериф. согл.	654	665	657	651	599	3226
V_j	3148	3149	3105	3151	2766	15319

$\chi^2 = 3{,}227$; $k = (5-1)(4-1) = 12$; $q_{i(k=12)}(\chi_q^2 > \chi^2) = 0{,}99$

Таблица 19

Текст Л. Кручковского
Текст Е. Шанявского

Класс	Выборка							
	I	II	III	IV	V	VI	VII	V_i
Гласные	1399	1437	1518	1506	1469	1502	2082	10913
Несогласные	690	744	777	767	767	761	1035	5541
Периф. согл.	535	558	550	564	515	549	729	4000
Непериф. согл.	795	845	873	819	816	847	1140	6135
V_j	3419	3584	3718	3656	3567	3659	4986	26589

$\chi^2 = 8{,}08$; $k = (7-1)(4-1) = 18$; $q_{i(k=18)}(\chi_q^2 > \chi^2) = 0{,}975$

Таблица 20

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	1577	2196	2332	2555	2647	11307
Несогласные	852	1089	1235	1255	1403	5834
Периф. согл.	624	835	948	975	993	4375
Непериф. согл.	756	1076	1168	1427	1371	5798
V_j	3809	5196	5688	6212	6414	27314

$\chi^2 = 26{,}191$; $k = 12$; $q_{i(k=12)}(\chi_q^2 > \chi^2) = 0{,}01$

Таблица 21

Текст С. Мрожека

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	2018	2629	2818	3057	3563	14 085
Несогласные	1020	1341	1531	1590	1943	7 424
Периф. согл.	796	1029	1056	1267	1378	5 526
Непериф. согл.	1040	1428	1527	1645	1931	7 571
V_j	4874	6427	6931	7559	8815	34 606

$\chi^2 = 12{,}363$; $k = 12$; $q_{i(k=12)}(\chi_q^2 > \chi^2) = 0{,}35$

В этом эксперименте наблюдается существенное снижение значения q_i для текста Е. Шанявского. Однако, учитывая снижение величины относительных частот и увеличение количества сравниваемых разрядов, мы снижаем величину уровня значимости до 0,005 (по сравнению с 0,05 в предыдущем эксперименте). В этом случае текст Шанявского оказывается однородным, хотя согласие и слабое. Показательны результаты по объединенному тексту (табл. 22).

Таблица 22

Объединенный текст

Класс	Выборка				
	Я. Ивашкевич	Л. Кручковский	Е. Шанявский	С. Мрожек	V_i
Гласные	6 438	10 913	11 307	14 085	42 743
Несогласные	3 216	5 541	5 834	7 424	22 015
Периф. согл.	2 439	4 000	4 375	5 526	16 340
Непериф. согл.	3 226	6 135	5 798	7 571	22 730
V_j	15 319	26 589	27 314	34 606	103 828

$\chi^2 = 22{,}83$; $k = (4-1)(4-1) = 9$; $q_{i(k=9)}(\chi_q^2 > \chi^2) = 0{,}006$

Здесь значение q_i еще более понижается, согласие уменьшается и находится почти на пороге области критических значений.

С переходом к членению согласных фонем на классы мы видим, что еще отчетливее вырисовывается грань между связными текс-

тами, характеризующимися внутренним единством, и текстами несвязными. Для несвязных текстов однородность относительно частотного распределения классов фонем «гласные», «несогласные», «периферийные согласные», «непериферийные согласные» устанавливается с трудом.

По-видимому, абсолютной однородностью любой текст обладает лишь в отношении самых общих классов — «гласные», «согласные», «несогласные», что же касается периферийных и непериферийных согласных, то, хотя в рамках нашего критерия гипотеза об однородности и не опровергается, полученные цифры демонстрируют не столь сильное согласие; следовательно, уместнее говорить не об абсолютной однородности, а о ярко выраженной тенденции к однородности.

Четвертая стадия

Следующее дихотомическое членение предусматривает выделение среди периферийных согласных компактных *kǩgǧx* и некомпактных *pṕbb́fḟvv́*, а среди непериферийных согласных — непрерывных *sśšzźž* и прерывных *tdcćčzźǯ*. Классы согласных и несогласных остаются без изменения.

Таблица 23

Текст Я. Ивашкевича

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	1329	1310	1319	1322	1158	6438
Несогласные	665	660	657	669	565	3216
pṕbb́fḟvv́	301	317	268	303	266	1455
kǩgǧx	199	197	204	206	178	984
sśšzźž	316	312	297	295	294	1512
tdcćčzźǯ	338	353	360	356	307	1714
V_j	3148	3149	3105	3151	2766	15319

$\chi^2 = 8{,}514$; $k = (6-1)(5-1) = 20$; $q_{i\,(k=20)}(\chi_q^2 > \chi^2) = 0{,}99$

Обращает на себя внимание исключительно высокое значение q_i для повести Я. Ивашкевича. Оно отражает композиционную однородность 1-й главы, отсутствие резких переходов действия во времени и пространстве и соответствующий указанным стилистическим чертам простой в лексическом плане язык персонажей.

Значение q_i по-прежнему достаточно высокое, но ниже, чем в предыдущем эксперименте.

Таблица 24

Текст Л. Кручковского

Класс	Выборка							
	I	II	III	IV	V	VI	VII	V_i
Гласные	1399	1437	1518	1506	1469	1502	2082	10 913
Несогласные	690	722	777	767	767	761	1035	5 541
pṕbb́fvv́	317	347	364	388	343	373	467	2 599
kḱgǵx	218	211	186	176	172	176	262	1 401
sśšzźž	362	389	398	364	380	410	512	2 815
tdcćčzźž	433	456	475	455	436	437	628	3 329
V_j	3419	3584	3718	3656	3567	3659	4986	26 589

$\chi^2 = 28{,}594$; $k = (7-1)(6-1) = 30$; $q_{i\,(k=30)}\,(\chi_q^2 > \chi^2) = 0{,}55$

Таблица 25

Текст Е. Шанявского

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	1577	2196	2332	2555	2647	11 307
Несогласные	852	1089	1235	1255	1403	5 834
pṕbb́fvv́	383	544	620	617	669	2 833
kḱgǵx	241	291	328	358	324	1 542
sśšzźž	373	507	526	666	627	2 699
tdcćčzźž	383	569	642	761	744	3 099
V_j	3809	5196	5683	6212	6414	27 314

$\chi^2 = 37{,}198$; $k = 4 \times 5 = 20$; $q_{i\,(k=20)}\,(\chi_q^2 > \chi^2) = 0{,}01$

Уровень значимости q = 0,005, введенный нами на предыдущем этапе эксперимента, будет сохранен и на последующих стадиях. Соответственно все значения критерия, вычисленные для каждого из четырех текстов, оказываются в области допустимых значений; следовательно, гипотеза об однородности относительно частного распределения классов фонем «гласные», «несогласные», *pṕbb́fvv́*, *kḱgǵx*, *sśšzźž*, *tdcćčzźž* не отвергается.

Как и на предыдущей стадии, согласие слабее в случае текста Е. Шанявского.

Теперь рассмотрим частоты выделенных классов фонем в объединенном тексте (табл. 27).

Таблица 26

Текст С. Мрожека

Класс	Выборка					
	I	II	III	IV	V	V_i
Гласные	2018	2629	2818	3057	3563	14 085
Несогласные	1020	1341	1531	1590	1943	7 424
pṕbḃfv́	493	661	675	757	864	3 450
kǩgǵx	303	368	381	510	514	2 076
sśšzźž	500	655	700	786	892	3 533
tdcćčʒʒ́ǯ	540	773	827	859	1039	4 038
V_j	4874	6427	6931	7559	8815	34 606

$\chi^2 = 22{,}552;\ k = 20;\ q_{i\,(k=20)}(\chi_q^2 > \chi^2) = 0{,}30$

Таблица 27

Объединенный текст

Текст	Класс				
	Я. Ивашкевич	Л. Кручковский	Е. Шанявский	С. Мрожек	V_i
Гласные	6 438	10 913	11 307	14 085	42 743
Несогласные	3 216	5 541	5 834	7 424	22 015
pṕbḃfv́	1 455	2 599	2 833	3 450	10 337
kǩgǵx	984	1 401	1 542	2 076	6 003
sśšzźž	1 512	2 815	2 699	3 533	10 559
tdcćčʒʒ́ǯ	1 714	3 320	3 099	4 038	12 171
V_j	15 319	26 589	27 314	34 606	103 828

$\chi^2 = 72{,}42;\ k = (4-1)(6-1) = 15;\ q_{i\,(k=15)}(\chi_q^2 > \chi^2) = 0$

Здесь картина совершенно меняется. Вычисленное значение χ^2 оказывается значительно больше, чем самое большое табличное χ_q^2 при k = 15 (39,7). Таким образом, при переходе к более дробному членению классов периферийных и непериферийных согласных однородность для объединенного текста отвергается. Можно сформулировать следующий результат: любой связный или несвязный текст будет однородным относительно частот гласных, несогласных и согласных, а также относительно частот периферийных и непериферийных согласных. Для связных текстов

это верно и в том случае, если класс периферийных подвергается членению на классы компактных и некомпактных, а класс непериферийных — на классы прерывных и непрерывных. Последнее утверждение неверно для несвязных текстов достаточно большого объема, а следовательно, и для всей совокупности.

Естественно, что в ходе последовательных членений мы хотим дойти до такого этапа, когда и связные тексты будут неоднородны относительно определенного частотного распределения классов фонем. Такая ситуация достигается нами при следующем дихотомическом членении, когда класс несогласных делится на носовые *mmnń* и ртовые *rlu̯j*; класс *pṕbb́ff vv́* делится по признаку «непрерывность» на *pṕbb́* и *ff vv́*; по тому же признаку класс *kk̄gǵx* делится на *kk̄gǵ* и *x*. Класс непериферийных прерывных делится на тусклые *td* и яркие (или аффрикаты) *cćčz̄ǯ́ž̄*.

Не будем приводить из-за их громоздкости таблицы абсолютных частот для каждого класса и каждого текста. Получены следующие результаты:

Текст Я. Ивашкевича — $\chi^2 = 49{,}1$; $k = (5-1)(10-1) = 36$; $q_{i(k=36)}$ $(\chi_q^2 > \chi^2) = 0{,}07$

Текст Л. Кручковского — $\chi^2 = 91{,}4$; $k = (7-1)(10-1) = 54$; $q_{i(k=54)}$ $(\chi_q^2 > \chi^2) = 0{,}001$

Текст Е. Шанявского — $\chi^2 = 68{,}90$; $k = 36$; $q_{i(k=36)}$ $(\chi_q^2 > \chi^2) = 0{,}0005$

Текст С. Мрожека — $\chi^2 = 48{,}3$; $k = 36$; $q_{i(k=36)}$ $(\chi_q^2 > \chi^2) = 0{,}07$

Объединенный текст — $\chi^2 = 176{,}25$; $k = 27$; $q_{i(k=27)}$ $(\chi_q^2 > \chi^2) = 0$

При принятом уровне значимости $q = 0{,}005$ значения критерия в текстах Л. Кручковского, Е. Шанявского, а также в объединенном тексте не удовлетворяют гипотезе об однородности.

Таким образом, последнее членение, которое удовлетворяло этой гипотезе при рассмотрении связных текстов, это: «гласные», «несогласные», *pṕbb́ff vv́*, *kk̄gǵx*, *sśšzźž̄*, *tdcćčz̄ǯ́ž̄*. Дальнейшее дробление приводит к появлению неоднородности уже внутри одного связного единого текста.

Итак, нами в деталях прослежена структура возникновения неоднородности текстов относительно фонологических частот. Эта неоднородность возникает задолго до выделения отдельных фонем в процессе идентификации и соотносится с введением в ходе этого процесса таких признаков, как «непрерывность—прерывность» и «яркость—тусклость», т. е. признаков, релевантных не для всей совокупности фонем, как признаки «гласность», «несогласность» и «периферийность», а лишь для определенных подсистем фонем.

Таким образом, критерий однородности текстов относительно частот классов фонем дает четкое и однозначное деление инвентаря

признаков. Хотя этот критерий основывается на чисто статистических показателях, результаты его применения совпадают с содержательными, смысловыми соображениями и оказываются фонологически значимыми.

3. Результаты эксперимента по проверке однородности с помощью многомерного критерия χ^2 сходны и с результатами, полученными при применении критерия χ^2 для альтернативного признака. Этот критерий позволяет уточнить, за счет каких именно классов фонем получается неоднородность. Как и в предыдущем случае, критерий согласия был применен внутри каждого текста, а затем — для объединенного текста.

Ниже приводятся результаты эксперимента.

Уровень значимости внутри отдельного текста 0,05, а в объединенном тексте — 0,005 (см. табл. 28).

Как явствует из таблицы, можно выделить классы фонем с более и менее стабильной частотой. К классам, частота которых стабильна (т. е. для которых значение q_i внутри одного текста больше 0,05, а в объединенном тексте — больше 0,005), относятся: гласные, несогласные, периферийные согласные, непериферийные непрерывные *sśšzźž*, периферийные некомпактные прерывные *pṕbḃ* и несогласные носовые *mḿnń*. Непериферийные прерывные *tdcćčzźž*, а также периферийные компактные *kkgǵx* и периферийные некомпактные непрерывные *ffvv́* обнаруживают нестабильность частоты как внутри отдельных текстов, так и в объединенном тексте. Несогласные ртовые *rlu̯j* имеют стабильную частоту внутри отдельных текстов, однако в объединенном тексте значение q_i лежит вне области допустимых значений.

Среди классов со стабильной частотой выделяются классы гласных, несогласных и периферийных согласных. Тот факт, что их частота не противоречит гипотезе об однородности при проверке по критерию χ^2 для альтернативного признака, соответствует обстоятельству, что именно эти классы можно считать по-настоящему альтернативными, т. е. приблизительно равными по частоте, всеобъемлющими и взаимоисключающими. Таким образом, однородность исследованных текстов по отношению к частотному распределению классов гласных, несогласных, периферийных согласных и непериферийных согласных, выявленная в эксперименте с многомерным χ^2, подтверждается подлинно альтернативным дихотомическим характером указанных различительных признаков, составляющих базовую структуру, лежащую в основе фонологической системы.

Что же касается остальных классов, показавших стабильную частоту при проверке по критерию χ^2 альтернативного признака, то при последовательных дихотомических членениях более крупных классов на более мелкие один из двух классов, получающихся в результате членения, имеет нестабильную частоту, а другой —

Таблица 28

Однородность четырех текстов польского языка в отношении частот классов фонем

Классы фонем	Сумма четырех текстов		Я. Ивашкевич. «Девушка и голуби»		Л. Кручковский. Первый день свободы»		С. Мрожек Пять рассказов		Е. Шанявский Пять рассказов	
	p	q_i	p	q_i	p	q_i	p	q_i	p	q_i
Гласные *a e i o u ō ē*	0,412	0,84	0,420	0,975	0,410	0,83	0,407	0,76	0,414	0,75
Несогласные *m n ṁ ń l r ų j*	0,212	0,30	0,210	0,92	0,208	0,92	0,214	0,15	0,214	0,15
Несогласные носовые *m n ṁ ń*	0,108	0,01	0,100	0,88	0,111	0,10	0,109	0,10	0,107	0,77
Несогласные ртовые *r l ų j*	0,104	0,0005	0,110	0,99	0,097	0,20	0,106	0,07	0,107	0,25
Периферийные согласные *p ṕ b b́ v v́ f f́ k ḱ g ǵ x*	0,157	0,07	0,159	0,80	0,150	0,55	0,160	0,12	0,160	0,40
Непериферийные согласные *t d c ʒ ć ʒ́ č ǯ s ś š z ź ž*	0,219	0	0,211	0,87	0,231	0,92	0,219	0,85	0,212	0,003
Периферийные компактные *k ḱ g ǵ x*	0,058	0	0,064	0,995	0,053	0,008	0,060	0,02	0,056	0,15
Периферийные некомпактные *p ṕ b b́ f f́ v v́* .	0,099	0,30	0,095	0,50	0,098	0,60	0,100	0,78	0,104	0,47
Непериферийные непрерывные *s ś š z ź ž*	0,102	0,40	0,099	0,60	0,106	0,75	0,102	0,96	0,099	0,15
Непериферийные прерывные *t d c ʒ ć ʒ́ č ǯ*	0,117	0,001	0,112	0,90	0,125	0,94	0,117	0,43	0,113	0,005
Периферийные непрерывные *f f́ v v́*	0,048	0	0,046	0,001	0,045	0,49	0,050	0,65	0,052	0,35
Периферийные прерывные *p ṕ b b́*	0,051	0,35	0,049	0,10	0,052	0,58	0,050	0,79	0,052	0,50
Непериферийные прерывн. яркие *c ć č ʒ ʒ́ ǯ*	0,050	0	0,045	0,70	0,059	0,95	0,047	0,94	0,046	0,05
Непериферийные прерывн. неяркие *t d*	0,067	0,05	0,067	0,62	0,065	0,48	0,069	0,25	0,067	0,005

стабильную (пунктиром обозначены классы с нестабильной частотой):

По-видимому, подобное строение фонологической системы в терминах ее статистических характеристик обеспечивает сочетание устойчивости с гибкостью — два качества, необходимые для того, чтобы система могла адекватно передавать информацию. Прежде всего устойчива частота наиболее общих альтернативных признаков — гласности, несогласности и периферийности. Внутри несогласных более стабильна частота носовых, т. е. фонем, характеризующихся, во-первых, наличием данного признака и, во-вторых, большей близостью к согласным. Внутри периферийных нестабильна частота группы компактных *kǩgǧx*. Для этой группы характерна определенная неравномерность фонологических отношений — отсутствие звонкого коррелята у *x* и весьма ограниченная встречаемость палатализованных *ǩ* и *ǧ* (по сравнению, допустим, с *ṕ* и *b́*). Далее, нестабильна частота группы *ff́ vv́*. Здесь также имеем дело с нарушением общих тенденций: частота звонких *vv́* больше, чем частота глухих *ff́*, что противоречит схеме в остальных классах фонем.

Среди непериферийных нестабильна частота прерывных фонем *tdcć čžźǯ*, получающих минус по признаку «непрерывность» в матрице идентификации фонем. Эта нестабильность в основном объясняется нестабильностью аффрикат *cć čžźǯ*, в то время как группа *td* имеет стабильную частоту в объединенном тексте (q_i = = 0,05). Тот факт, что именно аффрикаты нестабильны по частоте, может объясняться их сложным строением: для идентификации аффрикат необходимо наибольшее количество различительных признаков.

Таким образом, мы получаем, что в группу консонантных фонологических классов, обладающих стабильной частотой, входят лишь те классы, которые симметричны по своей внутренней структуре (т. е. у каждого члена такого класса обнаруживается полноценный коррелят по большинству признаков), и члены которых распределены согласно правилу Ципфа о том, что маркированный элемент встречается реже немаркированного, т. е. классы *pṕbḃ* и *sśšzźž*. Класс *td* лежит на границе, поскольку он не имеет палатализованного соответствия (соответственно его частота ста-

бильна в объединенном тексте и нестабильна в тексте Е. Шанявского).

Это положение формулируется нами как гипотеза. По-видимому, в любом языке система фонем устроена таким образом, что не все признаки функционируют одинаково во всех фрагментах системы: для каких-то фонем может не существовать противопоставлений по данному признаку, некоторые фонемы могут функционировать лишь в крайне редких случаях и т. п. Интересно проверить стабильность частоты подобных фонем и фонологических классов, в которые они входят. Если окажется справедливым, что наиболее стабильной частотой обладают фонологические классы, устроенные «наиболее правильным» образом, но при этом встречающиеся не слишком редко и распределенные в соответствии с правилом Ципфа, то можно будет говорить о позитивной связи между внутренним строением системы и ее статистическими характеристиками.

В этом случае окажется, что выделение некоторых классов согласных с наиболее стабильной частотой послужит удобным способом типологической индексации языка. В самом деле, тот факт, что в польском языке к числу фонологических классов со стабильной частотой принадлежат гласные, согласные и несогласные, не является, по-видимому, специфическим именно для польского языка, а характеризует устройство фонологической системы вообще, и в этом смысле представляет некоторую универсальную характеристику. Но то, что среди согласных выделяется именно класс *sšžzž* как класс со стабильной частотой, характерно для польского языка. Для другого языка, возможно, будет выделена своя, наиболее стабильная в отношении частоты группа согласных фонем. Более или менее стабильное присутствие в текстах подобной «диагностической» группы определяет общий характер звукового потока на данном языке, позволяет выделить звучание данного языка как специфическое.

Итак, подытоживая результаты экспериментов по проверке однородности текстов относительно частот классов фонем, можно сделать следующие выводы:

1. Членение системы фонем на классы, обладающие наиболее стабильной частотой, совпадает с членением на естественные фонологические классы. Каждый такой класс идентифицируется признаком или пучком признаков, среди которых доминирующую роль играет признак, определяющий силлабическую функцию данного класса, а также признак места образования (для согласных). Признаки способа образования и тембровые признаки играют второстепенную роль и подчинены признаку места: классы, идентифицируемые лишь по признакам способа образования или по тембровым признакам, имеют нестабильную частоту.

2. Естественные фонологические классы имеют более стабильную величину частоты по сравнению с классами, образованными

как случайные совокупности фонем. Эта стабильность показывает, что подобный естественный класс функционирует не только как простая арифметическая совокупность фонем, но что существует определенный механизм, регулирующий частоту именно того признака (или пучка признаков), который идентифицирует данный класс и что, таким образом, данный класс выступает как отдельная дискретная единица фонологического уровня.

3. Указанный механизм предположительно функционирует таким образом, что прежде всего обеспечивается постоянная доля в текстах наиболее общих, альтернативных фонологических признаков, управляющих преимущественно силлабической структурой речевого потока: гласности, согласности и несогласности. Далее, среди остальных классов выделяются наиболее правильно, просто и последовательно устроенные консонантные классы, члены которых, во-первых, распределены по правилу Ципфа, а во-вторых, встречаются не слишком редко. Эти классы также обладают стабильной частотой. Каждому подобному классу имеется соответствие в виде классов, обладающих нестабильной частотой. Такие классы, по-видимому, характеризуются нарушениями дистрибуции, отсутствием коррелят для некоторых членов и т. п.

§ 4. Эксперимент по проверке однородности с помощью порядкового критерия Н. В. Смирнова

Как это неоднократно подчеркивалось в предыдущем разделе, критерий проверки χ^2 имеет дело с реальными числовыми значениями частот исследуемых объектов.

В терминах проверки равенства наблюденных частот удалось установить, что в разных выборках одни и те же фонемы могут встречаться со столь различной частотой, что ее нельзя возвести к общему теоретическому прототипу. С другой стороны, обнаружено, что частота некоторых субфонемных единиц — классов фонем — в различных выборках может считаться статистически совпадающей. Таким образом, реальная картина наблюденной встречаемости фонем складывается под влиянием двух противоположных факторов: тенденции одних различительных признаков, входящих в состав фонемы, к стабильной частоте и тенденции других различительных признаков к нестабильной частоте.

Ограничивается ли, однако, статистическая структура текста на фонемном уровне только реальными цифрами наблюденной встречаемости фонем? Критерий χ^2 позволяет сравнивать лишь индивидуальные частоты данной фонемы в различных выборках. Он не дает возможности выяснить картину соотношения частот разных фонем. А между тем интуитивно очевидно, что имеется определенная закономерность в аранжировке частот фонем. Эта закономерность, однако, выявляется лишь при более

пристальном рассмотрении статистических данных. Приведем в качестве примера относительные частоты соответствующих фонем в двух выборках: четвертом и пятом рассказах Е. Шанявского (табл. 29).

Здесь двух различных текстах сопоставляются одинаковые фонемы, и их частота сравнивается. Полученная картина является типичной: у большинства фонем частоты довольно близки (у некоторых даже совпадают $k = 0{,}026$), однако для некоторых фонем расхождения весьма значительные: *r* — 0,026 и 0,036, *l* — 0,022 и 0,015, *ć* — 0,015 и 0,010, *j* — 0,021 и 0,025. Эти расхождения и приводят к тому, что в терминах критерия χ^2 обнаруживается неоднородность относительно всего ряда частот фонем, а поскольку, как это было показано ранее, нестабильность частоты может наблюдаться у самых различных фонем, подобное сопоставление одинаковых фонем скорее затушевывает, чем раскрывает закономерность в аранжировке частот фонем, которая ощущается интуитивно.

Произведем перегруппировку относительных частот в обоих рядах: расположим эти частоты по рангам, т. е. в порядке убывания, и сопоставим друг с другом уже не одинаковые фонемы, но одинаковые ранги. Таким образом, сравниваются события, состоящие не в «выпадении» данной формы, а в «выпадении» данного номера в списке по убывающей частоте (одинаковые номера могут соответствовать различным фонемам).

При данном расположении частот закономерность их соотношения становится более очевидной: в обеих выборках имеется одинаковое соотношение между частотами, начиная с первого ранга и кончая последним. Именно это постоянное соотношение, или, говоря иначе, динамика убывания частоты (или ее возрастания), начиная от самой частой фонемы (соответственно наиболее редкой) и кончая самой редкой (соответственно наиболее частой) и определяет интуитивное ощущение того, что уже на фонемном уровне присутствует статистическая структура, организующая систему фонем.

При подобном рассмотрении структура частот отделяется от конкретных носителей частот и изучается как самостоятельный, независимый механизм.

Однако поскольку встречаемость фонем связана, помимо прочего, со стабильной частотой некоторых фонологических классов, место каждой фонемы в списке по убывающей частоте не может быть совершенно произвольным: первые места в этом списке обычно занимают гласные *e i a o* и согласные *t* и *n* (взаимный порядок этих первых по частоте фонем может меняться, ср. для рассказов Шанявского: *eaoint*, *eaoitn*, *eoaitn*, *eaotin*, *eaoitn*), очевидно, что фонемы *ǵ ź* или *ǵ* (беря наудачу) не будут самыми частыми, а будут тяготеть скорее к концу списка. Поэтому в случае частот фонем существование независимой жесткой сетки соотно-

Таблица 29

	e	*i*	*a*	*o*	*t*	*n*	*j*	*s*
4-й рассказ	0,101	0,085	0,091	0,087	0,042	0,037	0,021	0,029
5-й рассказ	0,111	0,081	0,092	0,088	0,047	0,039	0,025	0,031

Продолжение

	r	*u*	*l'*	*k*	*p*	*z*	*m*	*ń*
4-й рассказ	0,026	0,036	0,022	0,026	0,032	0,020	0,033	0,025
5-й рассказ	0,036	0,034	0,015	0,026	0,033	0,017	0,037	0,027

Продолжение

	v	*d*	*c*	*č*	*g*	*ɣ*	*ś*	*v́*
4-й рассказ	0,024	0,029	0,010	0,014	0,013	0,030	0,018	0,010
5-й рассказ	0,025	0,026	0,012	0,012	0,010	0,030	0,014	0,011

Продолжение

	f	*ć*	*š*	*ž*	*x*	*j́*	*õ*	*ḱ*	*b*
4-й рассказ	0,013	0,015	0,022	0,016	0,011	0,002	0,010	0,007	0,012
5-й рассказ	0,017	0,010	0,020	0,014	0,010	0,002	0,005	0,005	0,010

Окончание

	ḿ	*ʒ́*	*ǵ*	*ʒ*	*b́*	*ź*	*ẽ*	*ṕ*	*ǯ*
4-й рассказ	0,008	0,008	0,001	0,001	0,002	0,002	0,001	0,004	—
5-й рассказ	0,009	0,008	—	0,001	0,003	0,002	0,001	0,003	—

Таблица 30

	1 ранг	2 ранг	3 ранг	4 ранг	5 ранг	6 ранг	7 ранг	8 ранг	9 ранг
4-й рассказ	0,101	0,091	0,087	0,085	0,042	0,037	0,036	0,033	0,032
5-й рассказ	0,111	0,092	0,088	0,081	0,047	0,039	0,037	0,036	0,034

Продолжение

	10 ранг	11 ранг	12 ранг	13 ранг	14 ранг	15 ранг	16 ранг	17 ранг	18 ранг
4-й рассказ	0,030	0,029	0,029	0,026	0,026	0,025	0,024	0,022	0,022
5-й рассказ	0,033	0,031	0,030	0,027	0,026	0,026	0,025	0,025	0,021

Продолжение

	19 ранг	20 ранг	21 ранг	22 ранг	23 ранг	24 ранг	25 ранг	26 ранг
4-й рассказ	0,021	0,020	0,018	0,016	0,015	0,014	0,013	0,013
5-й рассказ	0,017	0,017	0,015	0,014	0,014	0,012	0,012	0,011

Продолжение

	27 ранг	28 ранг	29 ранг	30 ранг	31 ранг	32 ранг	33 ранг	34 ранг
4-й рассказ	0,012	0,011	0,010	0,010	0,010	0,008	0,008	0,007
5-й рассказ	0,010	0,010	0,010	0,010	0,009	0,008	0,005	0,005

Окончание

	35 ранг	36 ранг	37 ранг	38 ранг	39 ранг	40 ранг	41 ранг	42 ранг
4-й рассказ	0,004	0,002	0,002	0,002	0,001	0,001	0,001	—
5-й рассказ	0,003	0,003	0,002	0,002	0,001	0,001	—	—

шения частот не сразу обращает на себя внимание; кажется, что величина частоты однозначно связана с данной фонемой.

Нам представляется, что только существование подобной сетки частот, независимой от носителей частот и накладываемой на них извне, может объяснить упорное стремление исследователей обнаружить статистическую стабильность на фонемном уровне. Поскольку для фонем такая сетка каким-то образом связана с характером самих фонем, происходит естественное смещение объектов, и стабильность начинают искать там, где ее нет, т. е. на уровне частоты встречаемости отдельных фонем.

Приведем следующую физическую аналогию, чтобы показать возможность независимого существования сетки частот. В статистике часто пишут о модели урн, в которой априори заложена возможность неравновероятного исхода. Если теперь предположить, что опыт проводится таким образом, что имеется n ящиков разного объема K_1, K_2. . . K_n, по которым может распределяться M объектов, принадлежащих n различным классам, причем так, что в каждый ящик попадают объекты лишь одного класса, и если устроить так, что в ходе каждого опыта в каждый ящик попадают объекты нового класса, скажем, в результате 1-го опыта получилось распределение K_1 (a), K_2 (b),..., K_n (n), где a, b, . . . n — ярлыки классов, а в результате i-ого опыта — распределение K_1 (l), K_2(a),..., K_n(j), где l, a,. . ., j — также ярлыки классов, то мы получим два рода данных: данные о частоте каждого класса, которые будут отличаться от опыта к опыту, и данные об объеме ящиков K_1, K_2. . . K_n, которые во всех опытах будут постоянны. Продолжая эту аналогию, можно сказать, что в наших экспериментах постоянным будет соответствие частоты определенного порядка определенному рангу.

Как известно, впервые на закономерность распределения частот в зависимости от ранга указал Дж. Ципф. Мы уже отмечали в предыдущих главах нашей работы, что так называемый закон Ципфа об обратной зависимости частоты от ранга, модифицированный впоследствии Джузом, Кутсудасом и Мандельбротом, оказывается в этой форме (равно как и в модификациях) не состоятельным. Точнее говоря, было найдено (в работе Р. М. Фрумкиной о статистическом изучении лексики), что ципфовская зависимость выполняется лишь для интервала слов от $r > 50$ до $r < 1\,500$.

Нам представляется, что имеет смысл взглянуть на закон Ципфа шире — не с его аналитической стороны, а в более содержательном плане. Несоблюдение ципфовой закономерности на лексическом уровне, с одной стороны, отражает сложное строение этого уровня в статистическом плане (здесь можно грубо выделить три совершенно различных статистических слоя — частые слова, слова средней частоты и слова, встречающиеся 1—2 раза); с другой стороны, оно показывает, что, возможно, существуют другие единицы этого уровня, для которых постоянство «частотной

сетки» (т. е. соответствие частоты определенного порядка определенному месту в списке этих частот в порядке убывания) будет справедливо.

Более явственно единообразие частотной структуры (т. е. соотношение частот в зависимости от порядка их убывания) должно проявиться на материале фонем. Это связано, в частности, с тем, что число разных фонем значительно ниже, чем число разных слов в тексте, поэтому распределение фонемных частот в зависимости от порядка их убывания будет обозримо невооруженным глазом. Кроме того, отметим еще один момент: фонем не только меньше, чем слов, но в любом тексте достаточной длины выступают все фонемы данного языка, в то время как совершенно ясно, что даже в очень длинных текстах будут выступать далеко не все единицы лексического уровня данного языка. Иными словами, набор всех фонем данного языка является в общем обязательным конституирующим моментом текста (исключения лишь подчеркивают правило), а слова не являются минимальными единицами такого рода. Это различие принципиально, и из него выводится тот факт, что зависимость Ципфа априори должна соблюдаться для фонем и не соблюдаться для слов. Для лексического уровня необходимо найти единицы, подобные фонемам, тогда можно будет говорить о строгом постоянстве частотной сетки и на этом уровне. Отметим, впрочем, что и на уровне словаря иногда удается обнаружить постоянство частотной структуры: в одной из наших статей [44] мы рассматривали вопрос о постоянстве частотной структуры словаря для некоторых английских текстов. Там было установлено, что некоторые тексты обладают одинаковым соотношением частых и редких слов, т. е. удалось показать, что постоянная частотная структура может существовать и в отвлечении от конкретных носителей частоты. В применении к словарю доказать подобный факт, конечно, труднее, чем относительно фонем: диапазон изменения частот у слов гораздо емче, чем у фонем, поэтому для совпадения частотных структур требуется определенная семантическая близость текстов. Из нетривиальных результатов, полученных в этой работе, упомянем факт совпадения частотных структур романа Дж. Джойса «Улисс» и частотного словаря английского письменного языка, составленного Дьюи. Дело в том, что частотная структура частотного словаря заведомо должна отличаться от частотной структуры целостного единого текста.

Поскольку в частотный словарь включаются слова из возможно более разнообразных выборок, в нем должно быть гораздо больше редких слов, чем в целостном тексте. Тот факт, что роман Джойса по своей частотной структуре далек от целостного тек-

---

[44] Д. М. Сегал. Некоторые уточнения вероятностей модели Ципфа. «Машинный перевод и прикладная лингвистика», 1960, № 5.

ста, дает дополнительные объективные характеристики его оригинального художественного метода (словотворчество, привлечение слов из других языков, нарушение правил построения связного текста и т. п.). Таким образом, существование независимой частотной структуры оказывается вполне реальной вещью.

Трудности начинаются там, где встает вопрос аналитического описания этой структуры так, чтобы ее различные виды могли быть отделены друг от друга. Это — прежде всего трудности, связанные с тем, чтобы цифры относительной частоты неизбежно являлись округленными, и поэтому возможны случаи равенства частот у нескольких лингвистических объектов или, иначе говоря, случаи нарушения непрерывности распределения. Для словаря эта проблема вырастает до громадных размеров, поскольку в больших текстах есть целые группы слов с одинаковой частотой. В случае фонемных распределений вопрос стоит не столь серьезно, однако, как это видно из приведенной таблицы, и здесь возникают трудности, связанные с тем, как приписывать ранг одинаковым частотам. Б. Л. ван дер Варден уделяет этому вопросу раздел своей книги [45]. Он пишет: «До сих пор мы предполагали, что x и y обладают непрерывными функциями распределения и отсюда следовало, что возможность осуществления события $x_i = y_k$ можно не принимать в расчет. Однако на практике x_i и y_k всегда представляются округленными числами и, следовательно, имеют дискретное распределение; поэтому вполне возможен случай, когда $x_i = y_k$. Спрашивается, как в таком случае нужно определять порядковые номера r_i и s_k.

...Были предложены различные методы. Например, для того, чтобы решить, какую из двух равных величин x_i и y_k считать большей, можно бросать монету. Можно также условиться приписывать средний порядковый номер $r = {}^1/_2$ тем равным величинам $x_i = y_k$, которые в случае их неравенства должны были бы иметь порядковые номера r и r + 1».

Мы в наших экспериментах использовали метод монеты, поскольку количество равных величин было сравнительно невелико. Однако для словаря проблема должна быть решена иначе. По-видимому, при рассмотрении частотной структуры текста, имеющего словник в десятки тысяч слов, различным рангам следует приписывать разные «веса»: для самых частых 100 слов следует рассматривать каждый ранг, а затем (до слова с r = 2000) каждый пятидесятый, или, может быть, сотый ранг, далее — каждый тысячный ранг. В этом случае частотная структура будет несомненно сглаженной, идеализированной, однако ее построение будет облегчено, а проблема слов с одинаковой частотой будет не столь трудной. Можно, с другой стороны, предложить

---

[45] Б. Л. ван дер Варден. Математическая статистика, стр. 353—354.

такой способ представления частотной структуры, при котором для большого количества слов с одинаковой частотой вычисляется так называемый средний ранг, т. е. если слова с r_i по r_j имеют одинаковую частоту, в качестве репрезентанта этой группы выбирается ранг $\frac{r_i + r_j}{2}$.

Вернемся, однако, к рассмотрению частот фонем, где подобных проблем не возникает. Мы не будем ставить своей задачей нахождение аналитического выражения динамики изменения частоты от самой частой к самой редкой фонеме. Существенным является вопрос, совпадают ли частотные структуры фонем в различных выборках. Даже на глаз рассмотрение рядов фонем по убывающей частоте обнаруживает сходство. Существует метод установления такого совпадения. Этот метод отличен от критерия χ^2. Могли ли мы применить критерий χ^2 для установления совпадения частотных структур? Нет, ибо критерий χ^2 является для этой задачи слишком сильным. Вместо ответа на этот вопрос он прежде всего дает ответ на вопрос о числовом равенстве частот, принадлежащих одному и тому же рангу. Ответ на этот вопрос нас в общем не интересует, поскольку он не представляет лингвистического интереса,— один ранг может принадлежать разным фонемам, к тому же постановка такого вопроса неправильна и с методической стороны: если нас интересует вопрос L, то мы должны применять критерий, дающий ответ на этот и только на этот вопрос, ибо критерий, дающий ответ и на другие вопросы, окажется ложным, так как приведет к отвержению H_0 в тех случаях, когда она заведомо верна. В практическом плане невозможность применения критерия χ^2 к сравнению распределений рангов была проверена нами при сопоставлении частотного распределения фонем, полученного на материале всего нашего текста, и частотного распределения польских фонем, полученного Марией Стеффен [46]. Подробнее мы будем говорить о работе М. Стеффен немного ниже, здесь же отметим, что на глаз частотные структуры обеих совокупностей представляются весьма близкими. В результате проверки критерием χ^2 получаем $\chi^2 = 103{,}8$ и $q_i = 0$. Таким образом, если бы мы стали применять для целей сравнения частотных структур критерий χ^2, мы снова получили бы результаты, похожие на уже имеющиеся. Эти результаты зависели бы от вида текста, его связности и т. п., а нам необходимо раскрыть явление, не зависящее от текста, не связанное с конкретными значениями частоты данных лингвистических элементов.

Следовательно, вместо критерия χ^2 нужно было найти в аппарате статистики новый критерий, дающий ответ именно на тот вопрос, который мы ставим: совпадают ли частотные структуры

---

[46] M. Steffen. Częstość występowania głosek w języku polskim.— BPTJ, 1957, zesz. 16.

двух текстов, не зависящие от конкретных значений встречаемости фонем.

Здесь мы решили обратиться к так называемым порядковым критериям, составляющим сравнительно новый раздел математической статистики. Вот как определяет порядковые критерии Б. Л. ван дер Варден: «Порядковыми критериями называют такие критерии, в которых используются не сами значения наблюденных величин, а лишь упорядоченность, т. е. соотношения $x < y$ и $x > y$ (между двумя измеренными величинами). Такие критерии не зависят от функций распределения случайных величин x и y и поэтому их называют независящими от распределения, или непараметрическими» [47].

Поскольку в непараметрической статистике используется лишь упорядоченность значений случайной величины, а не сами эти значения, порядковые критерии не могут дать ответа на вопрос о равенстве двух эмпирических вероятностей. Именно поэтому порядковые критерии весьма удобны для выявления закономерности в упорядочении частот.

Здесь встает вопрос о том, что считать значением случайной величины, которое будет использоваться для построения вариационного ряда, являющегося основой порядковых критериев. В лингвистике, в отличие от других областей применения математической статистики, мы имеем дело в основном с качественными случайными величинами. Это вносит существенные трудности, поскольку при работе с порядковыми критериями надо иметь данные не только о вариационном ряде, но и о частоте его членов. В случае словаря, где приходится оперировать большим количеством наблюденных частот, построение эмпирической функции распределения облегчается: значением случайной величины является частота данного слова, а частотою служит число, показывающее, сколько раз данная цифра частоты встречалась. Получаем ряд типа:

Частотность слова	2675	1876	935	...	50	35	20	...	3	2	1
Частота	1	1	1	...	10	26	35	...	2150	4156	8333

Распределение подобного типа характерно именно для лексической статистики и его называют частотным распределением. Такие распределения рассматривались Дж. Юлом и Г. Херданом.

При фонологической статистике построение аналогичного эмпирического частотного распределения невозможно, поскольку число наблюдаемых разрядов невелико — несколько десятков,— поэтому наблюденные цифры частоты не повторятся ни разу (за небольшим исключением). Каждая цифра частоты будет иметь

---

[47] Б. Л. ван дер В а р д е н. Указ. соч., стр. 321.

частоту 1 (иногда 2). Это обстоятельство, однако, помогает нам найти выход из создавшегося положения.

Будем считать, что совпадение относительной частоты фонем в двух выборках объясняется (как это и есть на самом деле) необходимостью иметь дело с округленными величинами. В таком случае получаем некоторое идеальное распределение, в котором каждая цифра частоты встречается всего один раз. Подчеркнем, что речь идет не о встречаемости фонемы, а о встречаемости данного значения частоты. Таким образом, получаем ряд типа:

Значение частоты	452/3148	321/3148	315/3148	291/3148	...	104/3148	103/3148	...	4/3148	2/3148	1/3148
Частота	1	1	1	1	...	1	1	...	1	1	1

Будем считать, что такого рода частотное распределение, в котором все значения частоты равновероятны (абсолютная частота каждого значения равна 1, а относительная 1/n, где n — число сравниваемых разрядов — фонем), присуще каждой из наших выборок и характеризует фонологический уровень в целом.

Тогда эмпирическое распределение будет строиться следующим образом: на оси абсцисс откладываются в порядке возрастания значения частоты. Наименьшему значению частоты будет соответствовать ордината, равная 1/n, следующему — ордината 2/n, следующему — ордината 3/n и т. п.

Последнему значению соответствует ордината n/n = 1. Когда несколько фонем имеют одну и ту же частоту в одной выборке или в двух сравниваемых выборках, каждый случай встречаемости такой частоты рассматривается, согласно предположению о виде распределения, как новая случайная величина, отличная от других. Порядок следования одинаковых значений определялся бросанием монеты. Таким образом, возможны ряды типа:

Частота фонемы	...	0,026	0,026	0,025	0,024	0,024	...
Частота	...	1	1	1	1	1	...

Подобное представление немного нарушает обычный способ построения частотных распределений, но не влияет на работу критерия, поскольку каждое новое значение функции распределения образуется как сумма предыдущих.

Итак, исходным материалом для сравнения служат две эмпирические функции распределения, построенные на вариационном ряде наблюденных значений частот фонем.

Задача сравнения двух эмпирических функций распределения, характеризующих две выборки, формулируется Б. Л. ван дер Варденом следующим образом: «Пусть результатами наблюдений

являются $L = m + n$ независимых случайных величин: $x_1 \ldots$, $\ldots, x_m$; $y_1, \ldots, y_n$, и пусть все x_i наблюдаются в одинаковых экспериментальных условиях, т. е. можно предположить, что все они имеют одинаковые функции распределения. Такое же предположение мы будем делать и относительно у. Допустим, что наблюдается некоторое различие эмпирических распределений х и у; например, все х могут оказаться больше, чем у, или область рассеяния х может быть шире области рассеяния у. Спрашивается, является ли различие эмпирических распределений следствием различия истиных распределений, или же оно чисто случайное?

Нулевая гипотеза H_0, подлежащая проверке, утверждает, что все х и у имеют одинаковые функции распределения и, значит, наблюдаемое различие эмпирических распределений является чисто случайным. Однако при этом мы не должны делать никаких специальных предположений о функции распределения х и у»[48].

Поскольку эмпирические функции, которые мы рассматриваем в нашем эксперименте, в качестве вариационного ряда имеют ряд фонемных частот, расположенный в порядке возрастания частоты, нулевая гипотеза, которая формулируется в ходе эксперимента, сводится к выяснению интересующего нас вопроса: одинакова ли динамика изменения наблюденных значений частот фонем в двух выборках.

При построении критерия для проверки гипотезы H_0 производится следующее преобразование. Берутся два эмпирических распределения $x_1, x_2, \ldots, x_m$ и $y_1, y_2, \ldots, y_n$ и из них составляется одно новое распределение так, что под первым порядковым номером идет наименьшее значение из всех х-ов и у-ов, далее идет следующее по величине значение х или у и так далее до L-ого, наибольшего значения. Получаем ряд типа:

$$x_1, y_1, x_2, y_2, x_3, y_3, \ldots \ldots x_{L-2}, y_{L-1}, y_L.$$

Согласно гипотезе h_0, все перестановки $L = m + n$ случайных величин х и у, образующих новое распределение, равновероятны. Таких перестановок имеется n!, следовательно, каждой из них соответствует вероятность 1/n!

Построение критерия для проверки гипотезы H_0 эквивалентно указанию критической области V, включающей в себя некоторые из n! перестановок. Если наблюденное расположение принадлежит области V, то гипотезу H_0 следует отвергнуть.

Проиллюстрируем на примере, какого рода перестановки х и у будут входить в критическую область и, следовательно, приведут к отвержению H_0. Допустим, у нас имеются следующие функции распределения с равновероятными значениями случайной величины (табл. 31).

---

[48] Б. Л. ван дер Варден. Указ. соч., стр. 325—326.

Таблица 31

Значение случайной величины	x	x	x	x	x	y	y	y	y	y
Значение функции $F(x)$	$\frac{2}{10}$	$\frac{4}{10}$	$\frac{6}{10}$	$\frac{8}{10}$	$\frac{10}{10}$	$\frac{10}{10}$	$\frac{10}{10}$	$\frac{10}{10}$	$\frac{10}{10}$	$\frac{10}{10}$
Значение функции $F(y)$	0	0	0	0	0	$\frac{2}{10}$	$\frac{4}{10}$	$\frac{6}{10}$	$\frac{8}{10}$	$\frac{10}{10}$

Как следует из определения эмпирической функции распределения, ее значение в каждой точке равно сумме всех накопленных значений ординаты. В данном примере мы видим, что все значения x больше, чем y. Соответственно, в новой функции распределение всех x будет идти перед y. Функция F (x) достигает своего наибольшего значения 1, в то время как F (y) еще равна нулю. Несомненно, что подобная перестановка должна обязательно лежать в критической области V.

Мерой различия сравниваемых функций распределения считается наибольшая по модулю разность между значениями функций распределения. В нашем примере разности функций распределения выстраиваются в следующий ряд:

| F (x) — F (y) | = 2/10, 4/10, 6/10, 8/10, 10/10, 8/10, 6/10, 4/10, 2/10, 0.

Здесь максимальное значение F(x) — F(y) равно 1, т. е. является максимальным теоретически возможным. Кроме того, разница здесь все время будет в пользу F(x), поэтому оговорка о том, что разность берется по модулю, не имеет значения. На практике, однако, возможны отклонения как в ту, так и в другую сторону, поэтому берется максимальное значение разницы, независимо от знака.

Естественно, что разница по модулю, равная единице, всегда будет находиться в критической области критерия. Где же крайнее значение разницы между двумя эмпирическими функциями, входящее в критическую область и образующее водораздел между нею и областью допустимых значений?

Установление этой границы производится с помощью так называемого критерия Смирнова, являющегося развитием более общего критерия Колмогорова. Критерий Колмогорова [49] служит для оценки расхождения экспериментальной функции распределения ее известным теоретическим прототипом. Критерий Смирнова применяется для установления факта принадлежности двух

[49] См. § 16 книги Б. Л. ван дер Вардена.

эмпирических функций распределения некоему общему непрерывному распределению (его характер нас не интересует). Задачей критерия Смирнова является нахождение вероятности того, что в двух случайных выборках максимальная разница экспериментальной функции будет больше или равна наблюденной разнице, обозначаемой $D_{m,n}$. Если эта вероятность меньше чем q, или равна q, где q есть заранее выбранный уровень значимости, то проверяемая посредством критерия гипотеза о том, что обе выборки взяты из одной и той же генеральной совокупности с непрерывной функцией распределения, отклоняется. Величины q такие же, как и в других критериях (например, в критерии χ^2): 0,1, 0,05, 0,01 и т. д.

Практическая работа с критерием Смирнова ведется следующим образом [50]. Строятся две эмпирические функции распределения. Вычисляется максимальная разница между ними по модулю, обозначаемая $D_{m,n}$. Затем находится специальная статистика

$$D_{m,n} \cdot \sqrt{\frac{m \cdot n}{m+n}},$$

где m и n — число наблюдаемых объектов соответственно в каждой эмпирической функции. Величина $D_{m,n} \cdot \sqrt{\frac{m \cdot n}{m+n}}$ обозначается символом u и распределена по функции Колмогорова, таблицы которой составлены и имеются в справочниках. Ищется то наименьшее табличное значение u, которое превосходит

$$D_{m,n} \cdot \sqrt{\frac{m \cdot n}{m+n}}.$$

Далее для этого табличного значения u находится соответствующая ему процентная точка k распределения Колмогорова. Величина (1—k) и дает искомое значение q_i. Соответственно, если u мало (а это бывает при малом значении разницы $D_{m,n}$), то мало и k. Тогда q_i получается достаточно большим и лежит выше уровня значимости равного 0,1; 0,05 или 0,01. Значение u для k = 0,95 (q = 0,05) равно 1,36, а для k = 0,99 (q = 0,01) равно 1,63.

Как видим, практическое пользование критерием Смирнова напоминает пользование критерием χ^2: сначала вычисляется наблюденное значение статистики, затем находится наименьшее табличное значение, превосходящее вычисленную статистику, и соответственно этому табличному значению, распределенному по известному теоретическому закону (распределение χ^2 или функция Колмогорова), находится процентная точка распределения, по которой определяется, лежит ли вычисленная статистика в облас-

[50] См. соответствующий раздел книги Я. Янко «Математико-статистические таблицы».

ти допустимых значений или в критической области. Разница в данном случае состоит в том, что в критерии χ^2 сразу находится значение q_i, а в критерии Смирнова сначала находится обратное ему значение k, а q_i определяется уже по этому значению.

Несомненным преимуществом критерия Смирнова по сравнению с критерием χ^2 является большая простота вычислений.

Насколько нам известно, настоящая работа является первым опытом применения критерия Смирнова к лингвостатистическим задачам. Представляется, что если точно сформулировать вопрос, на который критерий должен дать ответ, и обосновать соответствующие эмпирические распределения, этот критерий может быть с успехом применен для самых различных лингвистических объектов подсчета.

* * *

Приступая к проверке однородности двух выборок относительно частотного распределения фонем, мы приблизительно представляли, какого рода результаты следует ожидать. Эти результаты, по предположению, не должны были зависеть от вида текста и его размера. Тот факт, что критерий Смирнова применим только при сравнении двух выборок, представлял трудности такого же рода, что и сравнение двух выборок относительно частот фонем при помощи критерия χ^2. Следует ли перебрать все возможные сочетания по два из 22 выборок? Мы приняли решение прекратить эксперименты в случае получения достаточно высоких и устойчивых значений q_i, не дожидаясь перебора всех возможностей.

Всего было произведено 15 попарных сравнений на материале обследованных нами польских текстов. Результаты достаточно красноречиво свидетельствуют об однородности исследованных выборок относительно частотного распределения в порядке возрастания величины частоты.

Сначала мы сравнили друг с другом все пять рассказов Е. Шанявского. Результаты следующие.

I—II рассказы: $|F_I - F_{II}|_{max} = \frac{10}{82} = \frac{5}{41}$; $u = 0{,}5066$; $k = 0{,}043$; $q_i = 0{,}957$

I—III рассказы: $|F_I - F_{III}|_{max} = \frac{5}{41}$; $u = 0{,}55$; $k = 0{,}077$; $q_i = 0{,}923$

I—IV рассказы: $|F_I - F_{IV}|_{max} = \frac{3}{40}$; $u = 0{,}34$; $k = 0{,}0002$; $q_i = 0{,}9998$

I—V рассказы: $|F_I - F_V|_{max} = \frac{1}{8}$; $u = 0{,}56$; $k = 0{,}087$; $q_i = 0{,}913$

II—III рассказы: $|F_{II} - F_{III}|_{max} = \frac{4}{42}$; $u = 0{,}33$; $k = 0{,}00009$; $q_i = 0{,}99991$

II—IV рассказы: $|F_{II} - F_{IV}|_{max} = \frac{4}{42}$; $u = 0{,}43$; $k = 0{,}007$; $q_i = 0{,}993$

II—V рассказы: $|F_{II} - F_{V}|_{max} = \frac{4}{42}$; $u = 0{,}43$; $k = 0{,}007$; $q_i = 0{,}993$

III—IV рассказы: $|F_{III} - F_{IV}|_{max} = \frac{3}{41}$; $u = 0{,}33$; $k = 0{,}00009$; $q_i = 0{,}99991$

III—V рассказы: $|F_{III} - F_{V}|_{max} = \frac{4}{40}$; $u = 0{,}45$; $k = 0{,}0126$; $q_i = 0{,}9874$

IV—V рассказы: $|F_{IV} - F_{V}|_{max} = \frac{4}{41}$; $u = 0{,}44$; $k = 0{,}0097$; $q_i = 0{,}9903$

Все значения q_i настолько выше, чем уровень значимости 0,05, что невольно напрашивается вывод о том, что ситуация внутри одного языка заведомо удовлетворяет предположению об однородности двух выборок относительно частотного распределения фонем. Обычно столь высокие значения q_i в случае применения статистического критерия свидетельствуют о том, что на данном материале постановка данного вопроса тривиальна, ибо априори ясно, что ответ будет положительным. Следовательно, надлежит задавать подобный вопрос относительно более разнородного материала, ибо там будет существовать возможность неоднозначного ответа.

Получив столь высокие и последовательные значения q_i для рассказов Шанявского, мы решили взять наудачу еще несколько выборок из нашего материала. Результаты подтверждают мнение об однородности двух произвольных выборок из польских текстов относительно частотного распределения в порядке возрастания частоты.

Текст Л. Кручковского, 6—7-я выборки: $|F_6 - F_7|_{max} = \frac{8}{42}$; $u = 0{,}87$; $k = 0{,}549744$; $q_i = 0450256$

Текст Я. Ивашкевича, 1—2-я выборки: $|F_1 - F_2|_{max} = \frac{4}{42}$; $u = 0{,}44$; $k = 0{,}00973$; $q_i = 0{,}99027$

Текст Я. Ивашкевича, 5-я выборка — текст Л. Кручковского, 4-я выборка: $|F_5 - F_4|_{max} = \frac{5}{42}$; $u = 0{,}55$; $k = 0077183$; $q_i = 0{,}922817$

Текст Л. Кручковского, 1-я выборка — текст С. Мрожека, 5-я выборка: $|F_1 - F_5|_{max} = \frac{4}{41}$; $u = 0{,}44$; $k = 0{,}00973$; $q_i = 0{,}99027$

И, наконец, мы сравнили данные по всем текстам нашего эксперимента с результатами М. Стеффен: $| F_1 - F_2 |_{max} = \frac{3}{42}$; $u = 0{,}33$; $k = 0{,}00009$; $q_i = 0{,}99991$.

Величина q_i в этом случае весьма велика, что отражает вполне наглядную близость сравниваемых распределений и показывает уместность применения критерия Смирнова в противовес критерию χ^2.

Полученные результаты по проверке однородности двух выборок из одного языка относительно распределения фонемных частот в порядке их возрастания показывают, что в одном языке динамика изменения фонемных частот (от самой редкой к самой частой — или наоборот) одинакова в любой выборке и никак не зависит от ее объема. Иными словами, для любой выборки цифры частот фонем — от самой частой до самой редкой — достаточно близки (вне зависимости от того, какая фонема соответствует определенной частоте).

Как следует трактовать эти результаты?

С одной стороны, их можно считать экспериментальным подтверждением закона Ципфа, правда, в немного ином аспекте, чем это обычно формулируется.

Мы доказали, что распределения частот в порядке их возрастания, относящиеся к фонологическому уровню данного языка, практически идентичны. После этого в общем несущественно, какую аналитическую форму имеют эти распределения. Важно другое — со статистической точки зрения пропорциональное содержание редких и частых фонем (безразлично каких именно, существенны лишь редкость или частость) в любом тексте данного языка постоянно. Итак, самым принципиальным результатом настоящего исследования следует считать вскрытие двойной структуры статистического процесса, ведущего к синтезу реальных языковых текстов.

Непонимание того факта, что статистическая структура фонологического уровня не линейна, а состоит по крайней мере из двух взаимно независимых механизмов — механизма, синтезирующего частотную сетку, одинаковую для всех текстов и существующую в отвлечении от ее реального заполнения, и механизма, который довольно сложным образом в зависимости от конкретного сочетания признаков со статистически стабильной и статистически нестабильной частотой внутри данной фонемы приписывает эти частоты конкретным фонемам,— непонимание этого факта вело к совершенно ложным утверждениям о том, что частота данной конкретной фонемы будет всегда стабильна.

С другой стороны, можно было бы заметить, что вскрытие одинаковости частотного распределения фонем в порядке возрастания частот сводится к довольно тривиальному утверждению о том, что любой текст обязательно будет построен из

единиц фонологического уровня. Даже если бы все выводы, которые можно извлечь из подтверждения существования единой ципфовской зависимости на фонологическом уровне, сводились к этому одному выводу, его нельзя было бы считать полностью тривиальным. Действительно, тот факт, что распределения частот фонем в порядке их возрастания одинаковы, свидетельствует о том, что любой текст составлен из элементов данного фонемного инвентаря. Однако он кроме того свидетельствует о том, что на совершенно ином уровне — уровне статистической встречаемости — имеется столько же единиц (цифр частоты), сколько имеется фонем, а также о том, что подобно тому, как фонемы сохраняют свою идентичность в различных употреблениях от текста к тексту, функция распределения частот фонем также остается неизменной.

На деле обнаружение того, что самая частая фонема в любом тексте данного языка (в нашем случае польского) будет всегда иметь частоту одинакового порядка (в нашем случае — порядка 0,100), а также что это справедливо и в отношении всех других частот фонем, расположенных в порядке убывания (самая редкая фонема в польском имеет частоту порядка 0,0001), совсем нетривиально. В принципе множество вероятностей, равное 1 и ограничиваемое точками сверху и снизу 0,100 и 0,0001 (в принципе это могут быть и другие величины P_1 и P_n для другого языка), может быть разделено на 42 подмножества (в общем виде n — число фонем в данном языке) огромным количеством способов. Теоретически можно было бы себе представить, что из n фонем языка L две фонемы («гласные») занимают 50% текста, а остальные делят оставшиеся 50% так, что следующие две фонемы («сонанты») занимают 50% этих оставшихся 50%. Оставшаяся часть текста может делиться между «согласными» поровну. Или множество вероятностей может делиться таким образом, что половина набора фонем может образовывать 1% текста, в то время как остальные 99% делятся таким образом, что 90% приходится на одну фонему, а 9% — на остальные, и т. д. и т. п. Теоретически таких способов деления можно придумать бесконечно много.

Отсюда видно, что данное членение множества вероятностей для данного языка, т. е. данная частотная сетка (функция распределения частот) (даже в отвлечении от ее конкретного заполнения) — есть факт лингвистически релевантный и характеризующий именно данную совокупность. Именно лингвистическая релевантность того, что данный язык всегда будет сохранять определенное членение множества вероятностей на постоянные классы, количество которых определяется числом единиц (фонем) и ограничено сверху и снизу определенными величинами, позволяет говорить о лингвистической значимости факта установления постоянства частотной сетки и, следовательно, справедливости закона Ципфа.

По нашему мнению, тот факт, что частотное распределение сохраняется от текста к тексту независимо от того, какой фонеме приписывается конкретная частота, свидетельствует о статистическом тождестве данного языкового уровня для данного языка. Это еще один важный вывод, который можно извлечь из наших результатов. Статистическое тождество позволяет говорить о единой частотной структуре фонологического уровня. Обсуждая проблематику закона Ципфа на фонологическом уровне, нельзя не отметить наиболее существенного различия между фонологическим и лексическим уровнями, которое сразу бросается в глаза. На лексическом уровне частотная структура не является единой для данного языка. Этот факт общеизвестен. Однако попробуем взглянуть на него с немного иной точки зрения. Обычно в лексической статистике единицей подсчета служит словоформа или слово как совокупность парадигм. Таких единиц подсчета в достаточно крупных текстах насчитываются тысячи и десятки тысяч и множество этих единиц является принципиально открытым. Возможно, что на лексическом уровне структура устроена более иерархично, чем на фонологическом уровне: если вместе сгруппировать однокоренные слова или если вместе сгруппировать однокоренные слова плюс синонимы, или если подсчитывать не отдельные слова, а классы слов — семантические или лексико-грамматические и т. п., то выявится другая структура единиц лексического уровня, уменьшится число подсчитываемых объектов и обнаружится постоянная частотная структура соотношения более частых и более редких элементов. Далее, на фонологическом уровне удалось выделить небольшое количество весьма крупных классов, реальная частота которых стабильна. Возможно, что подобные же классы существуют и на уровне слов, хотя это утверждение и нуждается в экспериментальном доказательстве.

Во всяком случае, можно априори утверждать, что для любого языкового уровня верно следующее: статистическая структура нелинейна. Если она существует, то в виде независимой частотной структуры, на которую накладывается структура стабильных и нестабильных частот, приписанных некоторым весьма общим классам объектов данного уровня.

Подобное утверждение соотносится с обязательным фактом иерархического строения языка и иерархического строения каждого уровня. Обязательное выделение на каждом уровне элементарных различительных признаков, на пересечении которых образуются объекты данного уровня, соответствует расслоению структуры частот этих объектов на стабильные частоты одних признаков и нестабильные частоты других признаков.

Получив весьма высокие значения q_i при сравнении выборок из одного языка, мы решили привлечь материал из других языков, чтобы проверить различающую силу критерия Смирнова:

если бы столь же высокие значения q_i получались и при сравнении разных языков, это означало бы, что критерий слишком груб и не улавливает релевантных различий.

Были привлечены материалы по фонемным подсчетам для следующих языков, кроме польского, русского и чешского [51], английского [52], маратхи [53], шведского [54], болгарского [55].

Материалы для русского, чешского, английского и маратхи являются совершенно достоверными и надежными. Они основываются на обширных выборках, составленных из множества различных текстов. Материалы по шведскому языку являются результатом сведения вместе нескольких фонемных подсчетов, выполненных в разное время Г. Фантом, М. Вайсс и др. Каждый из отдельных подсчетов основывается на небольшой выборке, однако все вместе они могут дать представление о распределении «ранг—частота» для шведского языка. Особняком стоит фонемный подсчет по болгарскому языку — он выполнен на весьма небольшой выборке и не может служить надежным материалом. Однако мы решили привлечь и этот подсчет, чтобы иметь больше источников по славянским языкам.

Итого было взято семь языков, каждый из которых сравнивался со всеми другими. Таким образом, была исследована 21 пара распределений. Результаты см. в табл. 32 (приведены величины q_i).

Как явствует из таблицы, порядок величин q_i здесь принципиально отличается от того, что имело место при сравнении распределений «ранг—частота» внутри одного языка. Если в случае одного языка величины q_i имели порядок 0,9(!), причем их разброс был сравнительно невелик, то здесь наибольшее значение q_i достигает лишь 0,8220, а его наименьшее значение — 0,01273. Принимая в качестве границы значимости и незначимости q_i равное 0,05, получаем, что три значения q_i из 21—0,01273 (чешский—английский), 0,03545 (чешский — болгарский) и 0,01638 (русский — болгарский) свидетельствуют о принципиальном различии сравниваемых частотных структур и об отсутствии однородности. Четыре значения q_i показывают лишь слабое согласие: чешский — польский ($q_i = 0{,}0522$), русский — шведский ($q_i = 0{,}0522$), чешский — шведский ($q_i = 0{,}06809$), англий-

---

[51] H. K u č e r a. Entropy, redundancy and functional load in Russian and Czech. «American Contributions to the Vth International Congress of Slavists». The Hague, 1963.

[52] P. B. D e n e s. On the statistics of spoken English. «Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung», Bd 17, H. 1. Berlin, 1964.

[53] Shriram Vasudeo B h a g w a t. Phonemic frequencies in Marathi...

[54] M. W e i s s. Über die relative Häufigkeit der Phoneme des Schwedischen. «Statistical Methods in Linguistics». Stockholm, 1961, № 1.

[55] М. М а р и н о в а, Ас. М а р и н о в. Статистически изследвания на фонемите в българския книжовен език. «Български език», 1964, 2—3, стр. 172.

Таблица 32

Язык	Язык						
	Чешский 35 фонем	Маратхи 38 фонем	Русский 40 фонем	Польский 42 фонемы	Шведский 44 фонемы	Английский 44 фонемы	Болгарский 48 фонем
Чешский 35 фонем		0,1626	0,3657	0,0522	0,06809	0,01273	0,03545
Маратхи 38 фонем			0,2295	0,6270	0,4262	0,3657	0,1355
Русский 40 фонем				0,4502	0,0522	0,3399	0,01638
Польский 42 фонемы					0,6440	0,8220	0,2920
Шведский 44 фонемы						0,3154	0,1122
Английский 44 фонемы							0,06809
Болгарский 48 фонем							

ский — болгарский ($q_i = 0{,}06809$). Полученные результаты свидетельствуют о том, что критерий Смирнова может различать случаи близости и принципиального различия частотных структур фонологического уровня в различных языках.

В 15 случаях из 21 частотные структуры фонологического уровня достаточно близки: польский язык близок по распределению «ранг—частота» маратхи, русскому, шведскому, английскому и болгарскому; чешский язык близок языкам маратхи и русскому; русский (кроме указанных) — маратхи и английскому, маратхи (кроме указанных) — шведскому, английскому и болгарскому, шведский (кроме указанных) — английскому. Можно было бы предположить, что близость или различие частотных структур двух языков связано с количеством фонем в них: при близком (или равном) количестве фонем частотные структуры близки, а при достаточно различном — частотные структуры различаются. Тогда применение критерия Смирнова было бы излишним, ибо этот критерий сложным образом показывал бы тот простой факт, что в данных языках одинаковое (или неодинаковое) число фонем. Частотная структура во всех языках была бы объектом, производным от количества фонем, и не представляла бы самостоятельного интереса.

Однако это не так. Цифры, приведенные в таблице, показывают, что нет непосредственной и однозначной связи между близостью числа фонем и близостью частотных структур. Следовательно, при наличии достаточно большого числа сравниваемых языков близость (различие) частотных структур по критерию Смирнова может служить базой для некоторой типологии. Правда, пока неясно, какого рода результаты следует ожидать; наши данные показывают, что степень близости частотных структур, по-

видимому, не зависит от родства языков (польский ближе к английскому, чем к чешскому) или от близости фонологических структур (польский, в котором нет противопоставления гласных по долготе, а также дифтонгов, но развито противопоставление по палатализованности, ближе к английскому, в котором это противопоставление отсутствует, чем шведский, с которым английский обнаруживает значительную степень фонологического сходства).

Видимо, возможная типология будет отражать чисто статистические характеристики системы, нежели ее лингвистические особенности. Пока можно отметить, что близкородственные чешский и польский языки оказываются на разных концах шкалы: польский обнаруживает наибольшее согласие с гипотезой об однородности частотных структур фонологического уровня в разных языках, а чешский — наименьшее (болгарский вследствие некоторой ненадежности данных следует пока оставить в стороне).

Итак, оказывается, что частотные структуры фонологического уровня во многих языках близки. Этот факт делает еще более очевидным наш результат, утверждающий, что для одного языка соотношение цифр частот на фонемном уровне постоянно вне зависимости от текста, — поскольку эти структуры могут совпадать даже в различных языках, в одном языке это само собой разумеется.

С другой стороны, тот факт, что критерий Смирнова различает принципиально далекие частотные структуры, показывает правильность нашего рассуждения о том, что частотная структура является самостоятельным образованием, не зависящим ни от числа объектов, ни от их семантической близости — при почти равном числе фонем множество вероятностей может члениться самым различным способом.

Задача будущих исследований состоит в том, чтобы найти лингвистические или иные корреляты этого различия.

* * *

Подытоживая все сказанное в настоящей главе, в которой были изложены основные результаты экспериментов по проверке стабильности частот фонологического уровня с помощью двух различных статистических критериев: χ^2 и Смирнова, можно сформулировать следующее.

— До сих пор в фонологической статистике господствовало утверждение об одномерности статистической структуры на этом уровне. Отсюда вытекала интерпретация интуитивно ощущаемой закономерности в аранжировке частот фонем как требование и постулирование обязательной стабильности этих частот вне зависимости от типа выборки из общей совокупности. Это мнение, в частности, господствует в работах Г. Хердана.

— Наши эксперименты позволяют интерпретировать наблюдаемую закономерность в аранжировке частот фонем как видимый результат неодномерности статистической структуры. Эта неодномерность проявляется во взаимодействии следующих независимых факторов:

а) неизменной и независящей от выборки функции распределения частот, существующей независимо от конкретных фонем, которым эти частоты приписываются, и остающейся тождественной в пределах данного языка;

б) частотности конкретной фонемы как результата сочетания внутри одной единицы более элементарных конституэнтов — различительных признаков, один из которых (а именно, определяющий силлабическую структуру данного языка) имеет стабильную частоту в любом тексте, а другие — имеют тенденцию к нестабильной частоте, зависящей от типа текста.

Иными словами, статистический механизм неодномерен в двух смыслах: с одной стороны, стабильная сетка частот отделяется от конкретных фонем, с другой стороны, частота каждой фонемы — результат взаимодействия уровня стабильных частот признаков и уровня нестабильных частот признаков. Таким образом, статистический механизм образования высказываний на фонемном уровне обладает той же иерархичностью (double articulation), которая характерна и для языка вообще.

На статистическом уровне оправдывается положение о том, что наблюдаемые факты — результат действия сложных систем конструктов и что истолковать эти факты можно, лишь выявив стоящие за ними механизмы.

Глава четвертая

НЕКОТОРЫЕ ВЫВОДЫ И ПРАКТИЧЕСКИЕ ПРИЛОЖЕНИЯ

§ 1. Некоторые сравнения и выводы

1. В предыдущей главе уже упоминалось о том, что в 1957 г. в журнале Польского товарищества языковедов была опубликована статья Марии Стеффен, содержащая статистику польских фонем [1]. От более ранних подобных работ (см. цитированные во II главе книгу Д. К. Ципфа и статью Е. Ньюмена) статья М. Стеффен отличается прежде всего тем, что она ориентирована исключительно на описание фонологии польского языка, а не на сравнение языковых систем или на построение синтагматических моделей. Для подсчетов автором были выбраны 10 отрывков из текстов самого разнообразного содержания (от художественной литературы до научных статей) — каждый отрывок имел объем в 5000 фонем. М. Стеффен стремилась охватить все функциональные стили современного польского языка с тем, чтобы получить частотность фонем для языка в целом. В статье чувствуется стремление провести подсчеты на хорошем статистическом уровне; в частности, ее автор вполне отдает себе отчет в неоднородности различных привлекаемых источников в отношении частот фонем (ср., например, таблицу в тексте статьи М. Стеффен на стр. 147, где сравниваются частоты отдельных фонем для разных текстов и показывается различие этих частот). Поскольку целью М. Стеффен было не выделение фонем с более стабильной частотой и противопоставление их фонемам с менее стабильной частотой, а получение окончательных величин частоты фонем для всего польского языка в целом, она пытается преодолеть эту неоднородность последовательным усреднением частот фонем по мере включения в материал

[1] M. S t e f f e n. Częstość występowania głosek polskich. «Biuletyn Polskiego Towarzystwa Językoznawczego», 1957, XVI.

каждого нового текста. Таким образом, конечная величина частоты выглядит достаточно устойчивой. Выше мы подробно рассмотрели вопрос о стабильности частоты фонем. В свете нашего эксперимента представляется, что процедура, предлагаемая М. Стеффен, лишь затушевывает истинное положение вещей, заключающееся, по нашему мнению, в том, что говорить о статистической устойчивости частоты для всех фонем нельзя. Устойчивой является сетка частот фонологического уровня в данном языке, а прикрепление конкретных фонем к этой сетке определяется весьма сложной и неодномерной процедурой. То, что стабильность частот, которой, на первый взгляд, достигает М. Стеффен, является мнимой, выявляется из сравнения величин частот фонем, полученных в нашем подсчете и в подсчете польской исследовательницы.

Прежде чем привести таблицу соответствий укажем, что в двух сравниваемых подсчетах были избраны различные системы фонем. М. Стеффен опиралась на трактовку польской фонологии, принятую в работах К. Нитча, В. Яссема и др. Соответственно в ее работе палатализованные *ṕḿv́bf́* не считаются фонемами, а в записи текста между этими элементами и последующим гласным вводился *j*. Таким образом, основное различие сравниваемых результатов сводится к тому, что частота *j* у М. Стеффен значительно выше, чем у нас. Что касается установления частоты *ṕḿv́bf́*, то сделать это было легко, поскольку М. Стеффен провела статистику не только фонем, но и аллофонов. Следовательно, для сравнения двух подсчетов требовалось выделить в особые графы частоты палатализованных губных, объединить частоты *i* и *y*, а также скорректировать частоту *j* в соответствие с тем, что палатализованные губные при сравнении считались отдельными фонемами. Это было также нетрудно сделать: М. Стеффен указывает в статье, что если считать губные палатализованные отдельными фонемами, а *i* и *y* — одной фонемой, то общий объем обследованного текста составит не 50 000, а 49 214 фонем.

Отсюда делаем вывод, что частоту *j* надо уменьшить как раз на величину, на которую уменьшается общий объем текста. Получаем, что новая частота $j = 0{,}045 - \frac{(50\,000 - 49\,214)}{49\,214} = 0{,}045 - 0{,}012 = 0{,}033$. Частоты остальных фонем мы отнесли к новому объему выборки и получили частоты, которые можно непосредственно сравнивать с результатами нашего подсчета. Эти данные содержатся в табл. 1.

Для сравнения величин частот в обоих подсчетах мы избрали довольно грубую оценку — так называемое относительное отклонение, которое подсчитывается как отношение взятой по модулю разности сравниваемых относительных частот и средней между ними величины к этой средней величине: $\frac{|p - p|}{\bar{p}}$. Мы решили не проводить в этом случае сравнения по критерию χ^2 с целью выясне-

Таблица 1

Сравнение частот фонем, полученных в результате нашего эксперимента и в статье М. Стеффен

Фонема	Выборка автора	Выборка М. Стеффен	Среднее	Относительное отклонение, %	Фонема	Выборка автора	Выборка М. Стеффен	Среднее	Относительное отклонение, %
e	0,109	0,104	0,1065	2,35	*ć*	0,015	0,013	0,014	7,14
a	0,095	0,096	0,0955	0,5	*ž*	0,015	0,012	0,0135	11,11
o	0,088	0,088	0,088	—	*b*	0,014	0,012	0,013	7,77
i	0,081	0,086	0,0835	2,99	*g*	0,014	0,015	0,0145	3,45
t	0,044	0,044	0,044	—	*c*	0,013	0,015	0,014	7,14
n	0,040	0,041	0,0405	1,23	*f*	0,012	0,013	0,0125	4,0
m	0,034	0,028	0,031	9,68	*č*	0,012	0,012	0,012	—
u	0,032	0,035	0,0335	4,47	*ǵ*	0,010	0,010	0,010	—
p	0,031	0,028	0,0295	5,09	*x*	0,010	0,011	0,0105	4,76
r	0,030	0,036	0,033	9,09	*ź*	0,008	0,008	0,008	—
s	0,029	0,030	0,0295	2,04	*ḿ*	0,008	0,007	0,0075	6,66
k	0,027	0,027	0,027	—	*o*	0,006	0,006	0,006	—
ų	0,027	0,021	0,024	14,3	*ḱ*	0,006	0,007	0,0065	7,7
j	0,026	0,033	0,0295	11,9	*ǵ*	0,003	0,003	0,003	—
ń	0,026	0,026	0,026	—	*ṕ*	0,003	0,003	0,003	—
v	0,025	0,025	0,025	—	*ʒ*	0,002	0,002	0,002	—
d	0,023	0,023	0,023	—	*ź*	0,001	0,002	0,0015	33,33
l	0,021	0,021	0,021	—	*ǵ*	0,001	0,002	0,0015	33,33
š	0,021	0,020	0,0205	2,4	*ǵ*	0,001	0,001	0,001	—
ś	0,019	0,015	0,017	11,8	*ē*	0,001	0,001	0,001	—
z	0,017	0,018	0,0175	2,8		0,0001	0,0001	0,0001	—

ния устойчивости частоты фонем, поскольку ранее подобный тест был нами проведен (см. предыдущую главу, раздел о критерии Смирнова) и было установлено, что накопленная величина χ^2 значительно выше критической и что, следовательно, в терминах этого критерия обе выборки неоднородны. Данные о стабильности или нестабильности частот отдельных фонем, полученные в нашем эксперименте, значительно более надежны, чем те, которые можно получить, сравнивая лишь два массива — наш и М. Стеффен. Поэтому сравнительно простой способ сопоставления частот является здесь вполне удобным, дополняя то, что получено в ходе основного эксперимента. Показатель относительного отклонения не зависит от величины частоты и поэтому может быть применен как к большим, так и к малым частотам. Будем считать, что относительное отклонение больше чем 5% (включительно) достаточно, чтобы признать сравниваемые частоты существенно различающимися и не представляющими общей величины.

Рассмотрим, как делится все множество фонем в соответствии с величиной относительного отклонения. Образуются два класса: фонемы, для которых величина относительного отклонения меньше 5%, и фонемы, для которых эта величина больше или равна 5%. Большая часть фонем (29 из 42) имеет в обоих текстах достаточно близкую частоту — для них относительное отклонение меньше 5%. Естественно, что поскольку эта оценка грубее, чем по χ^2, по этому критерию больше фонем имеют стабильную частоту, чем по критерию χ^2.

При сравнении текста М. Стеффен и нашего текста состав класса с относительно устойчивой частотой получается довольно показательным. Сюда прежде всего входят все гласные, все переднеязычные согласные и несогласные (*t d s z n ń l*), а также заднеязычные согласные (*k g x*). Кроме того, в этот класс входят *v v́*, а также фонемы с малой частотой *ǵ ʒ́ ʒ́ z̃ ẽ ṕ b́* и некоторые другие согласные. Если мы сравним этот довольно обширный список с фонемами *s z o k a*, которые мы выделили после нашего эксперимента как достаточно стабильные в отношении частоты, то увидим, что все они входят в класс фонем со стабильной частотой, установленный при помощи исследования относительного отклонения.

Если мы теперь рассмотрим состав класса фонем с нестабильной частотой, то увидим, что этот класс включает губные согласные и несогласные (*m ḿ ṕ b*), часть шипящих (*ś ž ź*), а также плавные *r u̯ j*. В нашем эксперименте фонемы *r l u̯ j m x* выделялись как фонемы, имеющие тенденции к нестабильной частоте. Теперь мы видим, что и эти результаты в общем совпадают с полученными после сравнения нашего текста с текстом М. Стеффен (относительное отклонение для *x* равно 4,76% — формально *x* попадает в класс фонем со стабильной частотой, однако 4,76% слишком близко к 5%).

Таким образом, эксперимент, который можно считать до некоторой степени контрольным, подтвердил выводы, полученные нами в главе III, а именно, что определенная группа фонем (плавные + *x*) имеют ярко выраженную тенденцию к нестабильной частоте. Это связано, может быть, с особым положением плавных в системе фонем, а также с тем, что *x* является аномальным членом системы согласных — у него нет звонкого коррелята.

Результаты сравнения нашего подсчета с подсчетом М. Стеффен показывают, что система фонем устроена неодномерно по отношению к фактору частоты встречаемости. Поэтому стремление М. Стеффен найти процедуру для превращения нестабильной частоты в стабильную оказывается безуспешным.

2. Как уже указывалось в III главе, система фонем не является простым одномерным объединением одинаковых элементов, если рассматривать ее с точки зрения статистической встречаемости. В этом плане система фонем сходна с системой словаря. В словаре мы также наблюдаем разделение всего инвентаря на более ста-

бильную в отношении частоты часть и часть с менее стабильной частотой. Отметим, что по некоторым весьма предварительным данным (содержащимся, в частности, в статье Р. М. Фрумкиной «О законах распределения слов и классов слов» [2]) большей стабильностью встречаемости отличаются, с одной стороны, слова с большей величиной частоты (ср. гласные на фонологическом уровне), а с другой стороны — некоторые слова со средней частотой, употребление которых характеризует весь текст как совокупность (ср. фонемы *s z k* в польском языке на фонологическом уровне). Подобное сопоставление, разумеется, должно быть подтверждено дальнейшими разысканиями, однако уже сейчас представляется, что между статистической структурой встречаемости на уровне слов и фонем больше сходства, чем различия. Это сходство иногда проявляется не только в отношении стабильности—нестабильности частот, но и в отношении собственно статистической структуры, т. е. соотношения редких и частых элементов. Относительно словаря разделение на слова частые, со средней частотой и редкие стало общепринятым. Особенно это разделение стало актуальным после практических исследований, продемонстрировавших семантическое и функциональное различие указанных групп слов [3]. Рассмотрение таблиц частот польских фонем в порядке убывания показывает, что и на фонологическом уровне можно выделить три до некоторой степени сходных частотных класса: фонемы, употребляющиеся достаточно часто (в случае польского языка — это гласные *e a o i* с частотой примерно от 0,100 до 0,80), фонемы со средней частотой (до 0,015—0,014) и фонемы с малой частотой (до 0,0001). Верхняя частота групп фонем со средней и малой частотой примерно в два раза меньше нижней частоты предыдущей группы (нижняя частота у частых фонем $\approx 0{,}080$, а верхняя частота у фонем со средней частотой $\approx 0{,}040$; аналогично — нижняя частота у фонем со средней частотой $\approx 0{,}015$, а верхняя частота у редких фонем $\approx 0{,}008$).

Разумеется, характер подобного членения и состав классов зависит от языка. Например, сравнение имеющихся подсчетов по языкам английскому, шведскому, маратхи, чешскому, русскому и болгарскому (ср. гл. III) показывает, что не во всех языках членение на редкие, частые и средневстречающиеся фонемы столь четко, как в польском.

Наиболее четко это деление, кроме польского, соблюдается в русском языке и в маратхи: в обоих языках выделяются две самые частые фонемы (*а* и *і* в русском — частота их равна 0,12957 и 0,11351; *ā* и *a* в маратхи — соответственно 0,1436 и 0,1123).

Сб. «Структурно-типологические исследования». М., 1962.

Ср., в частности: P. G u i r a u d. Les caractères statistiques du vocabulaire. Paris, 1954, а также: Р. М. Ф р у м к и н а. Статистическое изучение лексики. М., 1965.

Частота следующей фонемы примерно в два раза меньше: 0,04629 в русском и 0,0683 в маратхи. В других сравниваемых языках ситуация немного иная. В чешском частоты пяти первых самых частых фонем: $e = 0{,}09648$; $o = 0{,}07785$; $a = 0{,}0699$; $i = 0{,}06425$; $s = 0{,}04994$ (небольшой скачок между e и o, а также между i и s; границу между частыми фонемами и остальным инвентарем следует провести, пожалуй, между i и s, поскольку частота s 0,04994 скорее тяготеет к области средних частот). В английском языке картина следующая: $a — 0{,}090445$; $t = 0{,}084033$; $i = 0{,}082537$; $n = 0{,}070849$; $s = 0{,}050893$. Здесь ситуация сходна с чешским языком (кстати частота самой частой фонемы в этих языках также близка), но границу провести легче: между n и s в частоте имеется заметная разница. Труднее всего границу между частыми фонемами и фонемами со средней встречаемостью установить для болгарского и шведского языков. В болгарском языке частоты самых частых фонем: $a = 0{,}0948$; $i = 0{,}0795$; $o = 0{,}0629$; $e = 0{,}0629$; $i = 0{,}0622$; $n = 0{,}0617$; $s = 0{,}0547$; $r = 0{,}0462$. Где провести границу? Разница между a и t составляет 0,0153, между t и o — 0,0166; разница между o и e, e и i, i и n незначительна; между n и s она равна 0,0070, а между s и $r = 0{,}0085$. В процентах к величине частоты эта разница составляет: между a и t — примерно 20% к частоте t; между t и o — примерно 26% к величине o и между s и r — примерно 18% к величине r. Из этого делаем вывод, что границу следует установить между t и o. В шведском языке, где последовательность частот — 0,082; 0,078; 0,067; 0,062; 0,053; 0,051; 0,049; 0,041, эту границу уже почти невозможно установить. Условно ее можно провести между 0,049 и 0,041.

Отметим, что чем выше частота самой частой фонемы в языке, тем четче отделяется группа фонем, имеющих высокую частоту. Соответственно, чем эта частота ниже, тем эту группу выделить труднее.

В заключение нашего беглого сравнения частотных структур фонологического уровня в разных языках укажем, что здесь открываются интересные возможности дальнейших исследований. Прежде всего, интересно выделить реальные параметры, а именно порядок величин частот для самой частотной и самой редкой фонемы в различных языках. Точные значения этих частот вряд ли удастся определить в силу возможной неоднородности, однако даже порядок этих величин может явиться интересной типологической константой языка, которую можно использовать при сравнении (мы видим, что диапазон частоты самой частой фонемы в исследованных нами языках достаточно велик — от 0,1436 в маратхи до 0,082 в шведском). Далее укажем, что не только сравнение величин частот терминальных фонем в списке по убывающим частотам может быть плодотворным; интерес представляет и изучение реального членения частотных структур на статистические подклассы, равно как и фонологического заполнения

этих подклассов. В процессе такого изучения можно использовать критерий Смирнова, дающий информацию о сходстве или несходстве функции распределения на всем диапазоне изменения частот. Отметим, что данные критерия Смирнова следует интерпретировать содержательно. В нашем случае (см. гл. III) чешский и английский языки, а также русский и болгарский оказались принципиально различными в свете критерия Смирнова. Если для русского и болгарского это различие частотных структур можно объяснить, вероятно, тем фактом, что в русском языке в отличие от болгарского четко выделяется группа из двух самых частых фонем, то для чешского и английского языков подобное объяснение неверно. Здесь дело в более тонких и не столь очевидных различиях частотных структур.

К сожалению, в настоящем исследовании мы не смогли заняться рассмотрением этого вопроса, следует предположить, что в этом направлении можно достичь существенных результатов.

Итак, подытоживая сравнение наших подсчетов частот польских фонем с соответствующими данными М. Стеффен, можно отметить, что основной вывод главы III о тенденции одних фонем к стабильной частоте, а других — к нестабильной частоте, подтверждается. Особенно четко выделилась группа фонем *r ų j m x*, для которой характерна нестабильность частоты.

3. Обратимся теперь к сравнению частот классов фонем в выборке М. Стеффен и нашей. Частоты фонем, подсчитанные М. Стеффен, были нами сведены в классы согласно процедуре, описанной в гл. III.

Таблица, показывающая результаты сравнения, приведена ниже (см. табл. 2).

Из таблицы явствует, что частота классов фонем в обоих подсчетах оказывается достаточно близкой. Однако не все классы имеют частоту, достаточно сходную для того, чтобы величина относительного отклонения оказалась ниже определенной заданной величины, которая в данном случае выбрана 3%.

Для плавных относительное отклонение равно 3,16%, для *x* — 4,76 и для периферийных некомпактных непериферийных *p ṕ b b́* 5,15%. Если для плавных большое значение относительно отклонения объясняется значительной нестабильностью каждой из фонем, входящих в этот класс (не происходит «компенсации»), то для *p ṕ b b́* такая нестабильность может объясняться тем, что в наших материалах, ориентирующихся на художественную литературу, моделирующую устное общение, больший удельный вес имеют слова «конативного» слоя коммуникации (Pan, Pani, prosze и т. п.).

Что же касается тех классов фонем, частота которых в результате нашего эксперимента была охарактеризована как стабильная или имеющая тенденцию к стабильности (гласные, несогласные, согласные периферийные и согласные непериферийные), то мы

Таблица 2

Сравнение частот классов фонем по данным автора и М. Стеффен

Классы фонем	Выборка автора	Выборка М. Стеффен	Среднее	Относит. отклон. в %
Гласные *a e i o u õ ẽ*	0,412	0,416	0,414	0,48
Несогласные *m ḿ n ń r ɫ l j*	0,212	0,213	0,2125	0,23
Плавные *r ɫ l j*	0,104	0,111	0,1075	3,16
Сонанты *m ḿ n ń*	0,108	0,102	0,105	2,86
Периферийные согласные *p ṕ b b́ k ḱ g ǵ x f f́ v v́*	0,157	0,157	0,157	—
Периферийные компактные *k ḱ g ǵ x*	0,058	0,061	0,0595	2,51
Периферийные компактные непрерывные *x*	0,010	0,011	0,0105	4,76
Периферийные компактные прерывные *k ḱ g ǵ*	0,048	0,050	0,049	2,04
Периферийные некомпактные *p ṕ b b́ f f́ v v́*	0,099	0,096	0,0975	1,54
Периферийные некомпактные непрерывные *f f́ v v́*	0,048	0,050	0,049	2,04
Периферийные некомпактные прерывные *p ṕ b b́*	0,051	0,046	0,0485	5,15
Непериферийные согласные *t d c z ć ź č ǯ s ś š z ź ž*	0,219	0,214	0,2165	1,15
Непериферийные непрерывные *s ś š z ź ž*	0,102	0,097	0,0995	2,51
Непериферийные прерывные *t d c z ć ź č ǯ*	0,117	0,117	0,117	—
Периферийные непрерывные яркие *c ć č z ź ǯ*	0,050	0,050	0,050	—
Периферийные непрерывные тусклые *t d*	0,067	0,067	0,067	—

видим, что для них величина относительного отклонения находится в пределах допустимых 3% (для гласных — 0,48, для несогласных—0,23%, для периферийных согласных эта величина составляет 0, поскольку частоты оказались совпадающими, для непериферийных согласных — 1,15%).

Таким образом, и для частот классов фонем результаты, полученные в наших экспериментах, оказываются подтвержденными.

4. Теперь, когда основные выводы, полученные нами в результате экспериментов по проверке стабильности частот фонем и классов фонем, подтвердились, попробуем показать, как осуществляется прикрепление отдельных фонем к стабильной сетке фонологических частот польского языка. Рассмотрим последова-

тельность фонем в порядке убывания частоты, полученную после усреднения данных наших подсчетов и подсчетов М. Стеффен: *e a o i t n u r m p s j k ń v ų d l š z ś g ć c ž b f č x ó ź ǵ k õ b́ ṕ ź ź f ǵ ẽ ž́*. Как мы уже отмечали выше, порядок закрепления фонем по отдельным местам в этом списке регулируется двумя закономерностями: стабильной частотой одних (немногих) классов и нестабильной частотой других. На это накладывается закономерность, известная под именем правила Ципфа и регулирующая не абсолютную встречаемость элемента, но его встречаемость относительно другого элемента, определяемого как маркированный (соответственно немаркированный) член оппозиции, в которую этот элемент входит вместе с первым: «из двух членов привативной оппозиции немаркированный член в связной речи встречается чаще, чем маркированный» [4].

Правило Ципфа и было избрано нами в качестве модели, объясняющей определенные закономерности в организации списка фонем по убывающей частоте. При этом наша трактовка ципфовской закономерности немного отличается от той, которую дает ей Н. С. Трубецкой, рассматривающий только привативные оппозиции. Во-первых, ципфовскую закономерность можно рассматривать в двух планах. С одной стороны, эта закономерность реализуется в виде безусловной тенденции фонологического класса, идентифицируемого минусовым значением данного признака, иметь меньшую частоту в тексте, чем класс, идентифицируемый плюсовым значением признака (мы переходим от трактовки фонологических признаков Н. С. Трубецким к дихотомическим различительным признакам Р. О. Якобсона). Эта тенденция соблюдается как в нашем материале, так и во всех других имеющихся подсчетах по самым различным языкам, и носит безусловный характер. С другой стороны, при переходе на уровень фонем мы отмечаем в нашем материале, что многие пары фонем, идентифицируемые соответственно плюсом и минусом одного и того же признака, также подчиняются ципфовской закономерности относительно частоты в тексте. Среди таких пар есть пары, объединенные строго привативными отношениями так, как их понимал Н. С. Трубецкой: «палатализованность—непалатализованность», «звонкость—глухость»,— но есть и такие, которые входят в более сложные отношения. Например, *t* противопоставлено с одной стороны *p*, а с другой — *k* как непериферийный — периферийному. Ясно, что, поскольку здесь одним отношением связаны три члена, это отношение нельзя считать привативным. Однако, если рассматривать каждую пару (*t* —*p* и *t* — *k*) в отдельности, то мы увидим, что отношение бинарное и что ципфовская зависимость выполняется.

То же самое можно сказать и об отношении «компактность—некомпактность» для *t* — *c*, *t* — *ć*, *s* — *š*, *s* — *ś*, а также об отно-

---

[4] Н. С. Трубецкой. Основы фонологии. М., 1960, стр. 292.

шении «яркий—тусклый» для $t - c$, $t - \check{c}$, $t - \acute{c}$ и для $d - z$, $d - \check{z}$, $d - \acute{z}$.

Как бинарные противопоставления, подчиняющиеся ципфовской зависимости, в нашем материале выступают также отношения «непрерывный—прерывный» для $t - s$, $d - z$, $p - f$, $k - x$ и отношения «назальный—неназальный» в системе гласных. При этом возможны случаи (они отмечены и в нашем материале, см. ниже), когда для некоторых бинарных отношений подобного рода ципфовская зависимость не выполняется. Это свидетельствует об особом положении элементов в системе, обусловливаемом факторами диахронии и т. п.

Таким образом, ципфовская зависимость выступает в нашей модели, с одной стороны, как безусловное преобладание частоты класса, идентифицируемого отсутствием признака, над частотою класса, идентифицируемого его присутствием, и, с другой стороны, как аналогичная тенденция для пар фонем, идентифицируемых общими бинарными различительными признаками. Для пар фонем эта тенденция не является безусловной. Следует отметить, что выявление бинарных отношений, подчиняющихся ципфовской зависимости, происходит внутри фонологических классов, организованных иерархически согласно порядку идентификации: *t* противопоставляется *k* не как некомпактный компактному, а как непериферийный периферийному, так как в нашей системе сначала выделяются классы периферийных—периферийных согласных, внутри которых уже происходит подразделение по признаку «компактность». То же самое справедливо и для признака «прерывность—непрерывность».

Итак, в основе нашей модели лежат, с одной стороны, безусловные закономерности, а с другой стороны, факты, эмпирически установленные в ходе анализа польского материала.

Рассмотрим подробнее, какие закономерности можно установить в расположении фонем в порядке убывания, руководствуясь правилом Ципфа в его вышеизложенной трактовке. Для этого прежде всего разделим весь список на зоны: частых фонем (*e a o i*, частота от 0,1065 до 0,0835); фонем со средней частотой (эта зона подразделяется на участок фонем с относительно высокой частотой: *t n u r m p s j k* — от 0,044 до 0,027 и участок фонем средней частоты: *ń v ų d l š ź ś g* — от 0,026 до 0,0145) и зону фонем с малой частотой (также подразделяющуюся на два участка: фонем низкой частоты: *ć c ž b f č x ó* — от 0,014 до 0,010 и фонем очень низкой частоты: *ź m̀ k̀ õ b́ ṕ ʒ ẑ f́ ǵ ẽ ž̃* — от 0,008 до 0,0001). Подобное подразделение во многом условно. Резкий скачок в частоте наблюдается лишь между самыми частыми фонемами и всем остальным списком. Членение в других зонах и участках осуществлялось не по частотному признаку, а по признаку фонологическому: мы стремились к тому, чтобы в каждую зону и подзону попадали фонемы, имеющие что-то общее в фонологическом плане.

Рассмотрим теперь каждую из зон в отдельности. В самое начало списка попадают четыре гласные фонемы *e a o i*. Они в польском языке всегда будут встречаться чаще, чем остальные фонемы, что отражает стабильность частоты класса гласных — их «более простое» в плане акустически-физиологическом устройство, чем устройство согласных. С другой стороны, гласные, обладающие дополнительным признаком назальности, — *õ* и *ẽ*, напротив, относятся к числу наиболее редких фонем. Эта закономерность прослеживается и в других языках, где система гласных включает некоторые дополнительные признаки: в чешском языке, где релевантный признак долготы, самая редкая фонема — *õ:*; в английском языке с развитой дифтонгизацией гласных наиболее редким элементом фонологической системы является дифтонг ээ. Что же касается порядка расположения фонем внутри группы, то он определяется следующими закономерностями: частота *e* как недиффузного больше частоты *i* как диффузного, частота *e* как непериферийного больше частоты *o* как периферийного. Эти закономерности выполняются в общем регулярно. Менее обязательной является закономерность: частота *e* как некомпактного больше частоты *a* как компактного.

Третья эмпирическая закономерность расположения фонем в первой группе — это то, что *a* превышает по частоте *o* и *i*. Здесь ципфовская зависимость неприменима, так как *a* не является бинарным коррелятом указанных фонем (оно как компактный противостоит всем остальным гласным в целом), однако эмпирический факт является вполне установленным. Таким образом, получаем два регулярных варианта порядка следования гласных этой группы: *e a o i*; *e a i o* и два менее вероятных, но вполне возможных: *a e o i*; *a e i o*.

Устройство группы самых частых гласных влияет на устройство группы следующих по частоте согласных и несогласных фонем только таким образом, что в эту последнюю группу попадает гласная фонема *u*, которая как диффузная и периферийная автоматически должна иметь меньшую частоту, чем гласные, имеющие минус либо по обоим, либо по одному из этих признаков. Группа из девяти фонем — *t n u r m p s j k* — может быть названа группой первичных согласных и несогласных. Сюда входят основные смычные *p t k*, далее разворачивающееся в сложные ряды звонких и палатальных согласных, а также аффрикат; *s*, далее дающий ряд свистящих и шипящих, плюс назальные *n m* и плавные *v j*. Таким образом, эта группа как в ядре содержит всю фонологическую систему польского языка. Порядок следования фонем в этой группе определяется свойствами самих этих фонем (это качество и входит в определение «первичные»). Здесь можно выделить следующие обязательные закономерности: частота *t* как смычного неперифериного глухого больше частоты *p* как смычного периферийного глухого; частота *p* как смычного периферийного некомпактно-

го глухого больше частоты *k* как смычного периферийного компактного глухого; частота *n* как назального непериферийного сонанта больше частоты *m* как назального периферийного сонанта; частота *t* как непериферийного прерывного согласного больше частоты *s* как непериферийного непрерывного согласного. Таким образом, в результате действия обязательных правил в этой группе устанавливается следующий порядок некоторых ее членов, являющийся костяком устройства этой группы: *t p k*; *n m*; *t s*. Пока эти три подгруппы не ориентированы в отношении друг друга (кроме того, остаются еще не определенные плавные *r j* и гласный *u*). Эту ориентацию можно условно произвести следующим образом: эмпирически устанавливается, что частота фонем *t* и *n* превосходит частоту фонем *r* и *j*, частота *n* превосходит частоту *s* (непрерывный сонант чаще непрерывного шипящего), а частота *u* как губного гласного превышает частоту *p* как губного согласного. Далее эмпирически мы устанавливаем, что частота *t* близка частоте *n*, и получим порядок: *t n . . . u . . . . . . p . . . k*, или порядок: *nt . . . u . . . p . . . k*, являющийся эмпирически менее частым. Фонема *s*, которая должна следовать после *t* и *n*, обычно следует и после *u*. Порядок же следования фонем *rjm* в указанных рамках свободный, т. е. мы можем иметь последовательности *tnmrups jk*, или *tnurmspjk*, или *tnrumpsjk*, или *tnumrspjk*, или *tnrjmupsk*, или *tnrujmspk* и т. д. и т. п.

Заметим, что первичные согласные и несогласные всегда присутствуют именно в этой части списка по убывающей частоте (правда, иногда в эту группу может входить фонема *ń*, но это не нарушает установленного порядка следования тех фонем, которые подчиняются этому порядку).

Тот факт, что *rmj* нельзя расположить относительно друг друга по правилу Ципфа (эти фонемы не являются членами бинарных оппозиций относительно друг друга), соответствует эмпирически зафиксированной нестабильности частот этих фонем. С другой стороны, если мы рассмотрим состав группы первичных негласных фонем, отличающейся, как было только что сказано, тем, что ее члены не могут иметь существенно более низкую частоту, то увидим, что в нее входит значительное число несогласных фонем *mnrj*. Присутствие здесь столь значительного числа несогласных, частота которых довольно высока, объясняет тот факт, что, несмотря на нестабильность частоты отдельных членов этого класса, его общая частотность все же является стабильной и составляет своего рода типологическую константу.

Далее начинаются зоны так называемых «вторичных» фонем, чье место в списке по убывающей частоте является производным от соответствующего места первичных фонем. Рассмотрим вторую группу фонем со средней частотой: *ńvụdlšzśg*. Сперва установим следующие зависимости: частота *ń* меньше частоты *n* (палатализованный — непалатализованный), частота *v* меньше частоты *p* (звон-

кий непрерывный — глухой прерывный); частота *d* меньше частоты *t* (звонкий прерывный переднеязычный — глухой прерывный переднеязычный); частота *š z* и *ś* меньше частоты *s* (непрерывный непериферийный компактный, непрерывный непериферийный звонкий и непрерывный непериферийный палатализованный — непрерывный непериферийный некомпактный, глухой, непалатализованный), наконец, частота *g* (заднеязычный смычный звонкий) меньше частоты *k* (заднеязычный смычный глухой). Помимо этого, частота любой из фонем *l ł* должна иметь меньше частоты любой из фонем *r j* (плавные непрерывные — плавные прерывные).

Все эти закономерности объясняют, почему рассматриваемые фонемы обязательно должны иметь частоту, меньшую, чем соответствующие первичные фонемы. Они, однако, не объясняют порядка расположения фонем внутри группы. Здесь приходится опираться на менее очевидные соотношения, чем в группе первичных согласных и несогласных. Прежде всего отметим, что любая из трех несогласных *ńlł* будет встречаться чаще, чем свистящие и шипящие *zšś*, поскольку сонант или плавный условно может считаться более просто устроенным, чем фрикативный согласный. Далее, *d* должно как непериферийный звонкий смычный встречаться чаще, чем периферийный *g*. Как прерывный непериферийный *d* будет встречаться чаще непрерывного *z*. Теперь рассмотрим положение фонемы *v*. Согласно правилу Ципфа, *v* как звонкая фонема должна была бы встречаться реже, чем ее глухой коррелят *f*. Этого, однако, не происходит: во всех славянских языках *v* занимает особое место в системе фонем, часто сближаясь по своим дистрибутивным характеристикам с сонантами, и встречается в речи значительно чаще, чем *f*. В наблюдаемой нами картине частотного распределения польских фонем особое место *v* проявляется не только в том, что его частота больше частоты *f*. Как периферийная звонкая непрерывная фонема *v* должна была бы встречаться реже, чем непериферийная звонкая непрерывная фонема *z*. Однако на практике *v* встречается чаще *z*.

И, наконец, еще одна зависимость, выведенная из расположения в группе первичных согласных и несогласных фонем. Как мы видим, в этой группе частота *s* превышает частоту *k*, соответственно в следующей группе частота *z* превышает частоту *g* (эта зависимость, впрочем, носит характер лишь тенденции). Таким образом, получаем костяк группы: (*ń l ł*) . . . (*š z ś*); *d* . . . *g*; *v* . . . *z*; *z* . . . *g*; *d* . . . *z*. Следовательно, можно вывести последовательность (*v d*) . . . *z* . . . *g*. С другой стороны, невозможно точно определить порядок следования фонем *zšś*. Здесь приходится констатировать, что язык допускает существенную свободу. Это свидетельствует, между прочим, о том, что фонему *ś* нельзя идентифицировать как палатализованный коррелят непалатализованной фонемы *š*; конечно, можно было бы выйти из этого положения, указав, что эта пара фонем также не подчиняется правилу

Ципфа. Однако, если для *v* — *f* такое положение имеет реальное фонологическое и диахроническое объяснение, то для *š* — *ś* такого объяснения нет. Следовательно, с точки зрения организации частотного распределения фонем, *ś* следует считать палатализованным коррелятом *s* (как это исторически и имело место), что вполне согласуется с правилом Ципфа (*s* в наших подсчетах встречается значительно чаще, чем *ś*).

Далее, невозможно четко определить взаимное расположение групп (*ń l ʒ́*) и (*v d*). Здесь также допускается существенная свобода.

Таким образом, общая схема имеет вид (*ń l ʒ́ v d*) . . . (*š z ś*) . . . *g*, фонемы в скобках могут группироваться произвольным образом. Введение дополнительных признаков — «палатализованности» и «звонкости», как мы видим, увеличивает свободу расположения фонем на «частотной лестнице»: остается меньше четко зафиксированных возможностей расположения фонем. Отметим, что подобная свобода связана с частотами среднего порядка (как это зафиксировано и в ходе наших экспериментов, см. гл. III).

Переходим к первой группе фонем малой частоты: *ć c ž b f č x ý*. Здесь мы попадаем в ситуацию, при которой наличие тех или иных фонем в группе целиком определяется предшествующими фонемами, в то время как внутренних закономерностей в расположении фонем в группе не удается установить. Прежде всего отметим, что интервал, внутри которого лежат частоты рассматриваемых восьми фонем, крайне невелик — от 0,014 до 0,010. Поэтому здесь свободное расположение фонем на «лестнице частот» незначительно отражается на величине самой частоты (в отличие от группы средних частот). То, что данные фонемы обязательно должны располагаться после фонем, помещенных в список первичных согласных и несогласных и следующий после него список фонем средней частоты, ясно из следующих закономерностей: частота любой из фонем *sśš* (неяркие непериферийные) будет обязательно выше частоты любой из ярких фонем *c ć č*; частота *š* будет выше частоты *ž* (глухой/звонкий), а частота *z* будет выше частоты *ž* как некомпактный versus компактный; частота *p* как глухого будет непременно выше частоты *b* как звонкого. Напротив, частота *v*, как уже указывалось, выше частоты *f*. Частота *k* должна быть больше частоты *x* (прерывный компактный периферийный — непрерывный компактный периферийный). Интересно, что поскольку *x* не имеет звонкого коррелята, его частота оказывается также меньше частоты звонкого *g*. Наконец, частота *ý* как палатализованного меньше частоты *v*. Отметим, что *v* нарушает еще одну закономерность: частота гоморганного фрикативного должна быть ниже частоты гоморганного смычного (ср. *d* — *z*). Напротив, частота *b* значительно ниже частоты *v*. Что же касается взаимного расположения *g* и *b*, то *b* должен был бы иметь частоту бóльшую, чем *g* (как некомпактный периферийный versus компактный периферийный). Однако на практике их места могут быть взаимозаменяемы (важно

лишь, чтобы g следовало перед x и после всех фонем, отмеченных как предшественники g в соответствующей группе). Таким образом, мы получаем группу фонем, расположенных в небольшом интервале частот, которые практически могут располагаться в этом интервале совершенно свободно, но обязательно после фонем предыдущей группы (случай g — b является здесь исключением, так как g формально относится к предыдущей группе). Отметим что в отличие от фонем двух предыдущих групп здесь свободное расположение в списке по убывающей частоте не отражается столь заметным образом на величине самой частоты, так как интервал очень невелик.

И, наконец, последняя группа фонем имеет очень малую частоту — от 0,008 до 0,0001. Состав этой группы более или менее постоянен, хотя отдельные ее члены, особенно имеющие более высокую частоту, могут перемещаться в предыдущую группу (например, $\acute{z}$ или $\acute{m}$), в то время как из этой группы некоторые фонемы могут спускаться вниз (например, f). Место каждой из входящих в последнюю в списке группу фонем определяется ее связями с фонемами, имеющими более высокую частоту. Прежде всего, выделим фонемы, чьи корреляты входят в предыдущую группу списка. Это $\acute{z}\acute{s}\check{z}\acute{f}\acute{b}$ (сюда же можно отнести фонемы $\acute{g}$ и $\acute{z}$, чьи корреляты либо формально входят в группу фонем средней частоты — z, либо могут быть помещены и в предыдущую группу — g, $\check{z}$). Соответственно указанные фонемы — $\acute{z}\acute{s}\check{z}\acute{f}\acute{b}\acute{g}\acute{z}$ должны иметь малую частоту и входить в самый конец списка по убывающей частоте. Среди этих фонем, в частности, $\acute{z}\acute{s}$ и $\check{z}$, идентифицируемые наибольшим количеством дифференциальных признаков,— наиболее «сложные» по своему составу. Остальные — $f\acute{b}\acute{g}\acute{z}$ — также обладают дополнительными признаками палатализованности и/или звонкости.

Помимо указанных фонем эта группа включает в себя фонемы $\acute{m}\acute{p}\acute{k}\tilde{o}\tilde{e}$, чьи немаркированные корреляты находятся среди частых фонем. Тот факт, что частота указанных фонем весьма невелика, вполне закономерен. Относительно назализированных гласных мы уже писали выше; что же касается периферийных $\acute{m}\acute{p}\acute{k}$, то их малая частота вполне соотносится с исследованным нами в главе III фактом уменьшения объема корреляции по палатализованности—непалатализованности в польском языке, в частности с отсутствием палатализованных периферийных в позиции конца слова. Таким образом, хотя формально эти фонемы могли бы выступать и в другом месте списка по убывающей частоте, фонологические причины заставляют эти элементы иметь крайне низкую частоту.

Внутреннее расположение фонем в последней группе задается следующими факторами. Во-первых, последнее место, как правило, занимают фонемы $\acute{z}$ и $\tilde{e}$. Эти фонемы являются маргинальными, выступают лишь в ограниченном количестве лексем и по существу

не могут считаться полноправными членами системы. Первые места обычно заняты фонемами $\acute{z}$ и $\acute{m}$. Здесь действуют следующие закономерности. Частота $\acute{z}$ (как палатализованного коррелята фонемы d) больше частот фонем z и $\check{z}$, ибо последние носят вторичный характер. Здесь особенно четко выявляется тот факт, что палатализованный яркий или шипящий по сути дела является палатализованным коррелятом соответствующего смычного или спиранта. Далее, частота $\acute{m}$ как палатализованного сонанта (несогласного) обычно больше частот $\acute{b}$ или $\acute{p}$ (палатализованных согласных); частота $\acute{k}$ (глухой) больше частоты $\acute{g}$ (звонкий); частота $\acute{b}$, как правило, больше частоты $\acute{g}$ (некомпактный смычный звонкий периферийный — компактный смычный звонкий периферийный); частота $\acute{p}$ (периферийный глухой смычный некомпактный) больше частоты $\acute{f}$ (периферийный глухой фрикативный некомпактный).

С другой стороны, частота $\acute{k}$ обычно больше частоты $\acute{p}$, что противоречит правилу Ципфа. Равным образом противоречит этому правилу и то, что частота $\bar{o}$ больше частоты $\bar{e}$. Частоты $\acute{b}$ и $\acute{p}$, как правило, равны.

Таким образом, получаем следующую последовательность: $(\acute{z}\acute{m}) \ldots (\bar{o}k) \ldots (\acute{b}\acute{p}\acute{z}\acute{z}) \ldots (\acute{f}\acute{g}\bar{e}) \ldots \check{z}$. Несколько бо́льшая организованность этой последовательности по сравнению с предыдущей связана с большей величиной частотного интервала, а также с тем, что в нее последовательно входят целые фонологические классы: палатализованные периферийные, звонкие аффрикаты, а также назализованные гласные.

Итак, нам удалось обнаружить определенный порядок «прикрепления» фонем к частотной сетке. Существенным представляется то, что этот порядок предусматривает сочетание обязательных жестких правил следования некоторых фонем друг за другом в порядке их частоты и свободного размещения других фонем. Основой этой модели послужило правило Ципфа, объяснившее как наличие правил следования фонем, так и отсутствие этих правил — фонемы, связанные в бинарные нейтрализуемые отношения, подчиняются правилу Ципфа и образуют «костяк» частотного распределения, а фонемы, не связанные такими отношениями и имеющие средние частоты, этому правилу не подчиняются и располагаются внутри соответствующей частотной зоны свободно. Здесь особенно показательны плавные фонемы $rl\eta j$, отношения между которыми не изоморфны отношениям между согласными фонемами. Эта модель хорошо иллюстрирует стабильность / нестабильность классов фонем — те классы, большинство представителей которых попадает в одну зону частотного списка, оказываются стабильными в смысле частоты (гласные — в 1-ю зону, большинство несогласных — во 2-ю зону).

Заканчивая наше описание модели прикрепления фонем к различным участкам «частотной лестницы» в зависимости от призна-

кового состава фонем и тех отношений, в которые они вступают с другими фонемами, следует подчеркнуть, что частотная сетка отнюдь не подразумевает неизменности конкретных частот ее составляющих. Для данного языка неизменна функция, руководящая изменением фонологических частот от самой большой до самой малой — так сказать, «наклон лестницы». Поэтому не следует полагать, что закрепленный порядок следования некоторых фонем в списке по убывающей частоте означает их постоянную относительную частоту. Отнюдь нет, ибо сами границы частотных зон, внутри которых этот порядок имеет место, могут в определенных пределах меняться.

Таким образом, в заключение можно еще раз отметить, что конкретная частота фонемы зависит от комплекса факторов, объединяющих факторы обязательно выполняемые и те, в которых проявляется свобода выбора. К обязательно выполняемым факторам относятся: постоянство функции распределения фонологических частот для данного языка, постоянство следования частот некоторых фонем друг за другом на некоторых участках функции (в зависимости от правила Ципфа) и стабильность частот наиболее общих классов фонем (гласные — несогласные — согласные). К факторам, объясняющим вариативность частот, относятся: нестабильность частот остальных фонологических классов плюс свобода размещения некоторых фонем внутри отдельных участков «частотной лестницы».

§ 2. Наблюдения над частотой фонологических классов в поэтических текстах (Ю. Тувим) как одно из практических приложений статистического анализа польской фонологии

Естественным выводом из предыдущих наблюдений следует считать возможность довольно большой вариативности частот тех фонологических классов, которые оказались нестабильными в частотном отношении. Коль скоро было показано, что различные консонантные признаки обладают в различных текстах разной встречаемостью, следует ожидать, что эта возможность где-то будет использована для целей языковой коммуникации. И здесь внимание обращается к поэтической речи. Общеизвестно, что именно в поэзии звуковая сторона может стать объектом целенаправленных изменений, может по сути дела представлять собою цель сообщения. Поэтому следует ожидать, что вариативность частоты фонологических классов резче выявится именно в поэзии. Для того, чтобы иметь возможность сравнивать данные о частоте фонологических классов в поэзии с некоторым значимым фоном, необходимо было принять определенный набор данных в качестве системы отсчета. Такой системой в настоящем разделе послужат

данные о частоте фонологических классов, усредненные по нашей работе и подсчетам М. Стеффен (см. табл. 2, стр. 198). Эти усредненные данные ни в коем случае нельзя понимать как общеязыковые частоты или даже частоты, характеризующие письменную форму польского языка. Это всего лишь некоторые усредненные данные, получившиеся по двум большим выборкам. Поскольку, как показало сравнение двух независимых подсчетов — нашего и М. Стеффен,— величины частот оказываются не очень сильно расходящимися, а в некоторых случаях близкими, можно с определенной долей достоверности предположить, что эти данные являются вероятными средними по польской прозе.

В качестве непосредственного объекта исследования в сфере польской поэтической речи мы решили избрать фонологическую структуру замечательного стихотворного цикла Ю. Тувима «Słopiewnie». Широко известны словообразовательные новшества, введенные Тувимом в этом цикле и призванные передать особый, «pierwotny» смысл вещей, который скрыт в словах, возродить дух язычества, «праславянства», присущий природе и языку. Этот цикл выделяется на фоне польской поэзии, в том числе и поэзии раннего Тувима, не только в этом, наиболее очевидном, плане. Звуковая фактура стихов «Słopiewnie» явственно воспринимается как существенно отличающаяся от того, что кажется привычным даже в стихах самого Тувима. Иначе говоря, можно заранее предположить, что встречаемость фонологических классов в этом цикле будет иной, чем в прозе, а также в других стихах [5].

Мы попытаемся сформулировать эти различия. Наши выводы будут носить достаточно предварительный характер, так как они основываются на анализе не всего корпуса стихотворений Тувима (что потребовало бы грандиозной технической работы), а лишь 32 стихотворений, из которых 7 непосредственно предшествуют циклу «Słopiewnie» в сборнике «Czwarty tom wierszy». Эти 7 стихотворений (Epos, Biologia, Świt, Wlesie, Zielona ziemia, Dziwy na niebe, Rzeź brzóz) образуют замкнутый цикл, как бы вводящий в поэтический мир «Słopiewnie». Сам цикл «Słopiewnie» включает в себя следующие 6 стихотворений: Zielone słowa, Słowisień, Kalinowe dwory, Wanda, O mowie rosyjskiej, Św. Franciszek. Остальные 25 выбраны наугад из всех предыдущих сборников поэта. Это стихотворения: Duma, Mędrkom, Szczęście, Czereśnie, Podmuch wiosny, Na noże, Właściwie, Litania, Colloquium, Humoreska, Piotr Plaksin (3 главы), Biały dom, Na ulicy, Rzuciłbym to wazystko, Niczyj, Siostro-wiosno, Krzyk, Dom, Ojczyzna, W barwistanie, Dziurawię niebo, Do generałów, Nauka. Общий объем просмотренного материала —17410 фонетических сегментов («Sło-

[5] J. T u w i m. Dzieła «Czwarty tom wierszy». Warszawa, 1955. Этот цикл относится к раннему творчеству Ю. Тувима (период до первой мировой войны).

piewnie»—882; 7 предшествующих стихотворений —2210, 25 остальных стихотворений —14318).

Нами были подсчитаны частоты фонем в каждом из обследованных стихотворений, а затем частоты фонем, образующих фонологические классы, идентифицируемые общими различительными признаками, были суммированы аналогично тому, как это делалось при проверке на однородность, но только подробнее: поскольку нас интересовали специфические особенности стиха, было выделено большее количество классов. Все множество фонем было разбито сначала на три крупных класса: гласные, согласные и несогласные, а затем каждый из этих классов членился на более дробные подклассы: гласные на гласные высокой тональности (*ieẽ*), низкой тональности (*oõu*) и компактный *a*; несогласные членились на носовые (*mm̓nn̓*) и ртовые (*rlu̯j*); согласные — на периферийные и периферийные; периферийные — на компактные (*kk̓gg̓x*) и некомпактные (*pp̓bb̓ff̓vv̓*); компактные — на прерывные (*kk̓gg̓*) и непрерывный (*x*); некомпактные аналогично — на прерывные (*pp̓bb̓*) и непрерывные (*ff̓vv̓*); непериферийные членятся на прерывные (*tdcʒć̓ʒ̓čǯ*) и непрерывные (*sśšzźž*), из которых первый подкласс в свою очередь делится на неаффрикаты *td* и аффрикаты *cć̓čǯʒ̓ʒ̓*. Помимо этого все согласные делились по трем общим признакам: звонкость—незвонкость, прерывность — непрерывность и яркость—неяркость (последняя пара признаков положительно идентифицирует согласные *sśšzźžcć̓čǯʒ̓ʒ̓*). Все согласные и несогласные делились также на два класса по признаку «палатализованность — непалатализованность». Поскольку здесь нас интересуют не классы, дающие максимально стабильную частоту, а как раз те классы, где может проявиться специфика поэтической обработки звуковой фактуры, мы исследовали и такие классы (палатализованные, звонкие, непрерывные), которые заведомо будут иметь нестабильную частоту — хотелось узнать, насколько велико будет отклонение от нашего материала.

Для того чтобы понять значимость полученных частот в стихотворениях «Słopiewnie» их, как отмечалось, следует сопоставить с некоторым «фоном». Разумеется, вследствие небольшого размера каждого стихотворения и сознательной ориентированности многих из них именно на план выражения следует заранее ожидать, что внутри одного стихотворения частоты классов фонем (и, следовательно, признаков, идентифицирующих эти классы) будут сильно отличаться от любого «среднего уровня». Так и получилось (даже частотность гласных в отдельных стихотворениях сильно отличается от того, что получилось в наших подсчетах: в стихотворении «Ojczyzna»—0,371, а в стихотворении «Zielone słowa»— цикл «Słopiewnie»—0,446; подобные отклонения указывают на значение величины текста; оба стихотворения крайне невелики по объему), поэтому мы будем рассматривать не частоты в каждом отдельном стихотворении, а картину для всего цикла

в целом. Фоном для оценки частот классов фонем в цикле «Słopiewnie» последовательно являются: средние частоты по польской прозе, частоты, полученные для 25 случайно выбранных стихотворений раннего Тувима, и частоты, полученные для 7 стихотворений, предшествующих циклу «Słopiewnie».

Прежде всего охарактеризуем частотность фонологических классов в 25 случайно подобранных стихотворениях Ю. Тувима. Средние частоты классов фонем, выведенные для 25 стихотворений, выбранных наугад, в общем, за исключением некоторых значимых случаев, близки к средним по прозе. Из расхождений нам представляются существенными следующие: понижение процента гласных в стихах по сравнению с прозой (из 25 стихотворений в 13 процент гласных ниже общеязыкового, в 9— совпадает с общеязыковым и лишь в 3— выше общеязыкового).

При этом стихотворения, в которых процент гласных выше наших средних результатов по прозе, во-первых, крайне невелики по объему и, во-вторых, в пределах этого небольшого объема сознательно ориентированы на звукопись гласными («Właściwie» и «Colloquium», особенно второе, представляют собою поэтическое моделирование устной речи с ее многочисленными вспомогательными словами типа a, i, no, i и т. п., дающими увеличение процента гласных. «Siostro-wiosno» — уже подступ к «Słopiewnie»— сознательное использование игры гласных).

В более «обычных» (т. е. ориентированных на польскую поэтическую традицию) стихотворениях раннего Тувима — Duma, Litania, Dom и других — мы наблюдаем последовательное понижение содержания гласных по сравнению с прозой. Это может быть связано с большей синтаксической «теснотой» стиха (в смысле Ю. Н. Тынянова): в просмотренных 25 стихотворениях гораздо меньше синтаксических связочных элементов (союзов, предлогов), обычно содержащих вокалические сегменты, чем в прозаическом тексте.

Отметим, что понижение содержания гласных в поэтических текстах отмечено и на материале украинского языка [6]. По данным подсчетов В. И. Перебейнос и ее группы частоты гласных следующие: 0,41423— поэзия; 0,41630— научная проза; 0,41658— публицистика; 0,42157— устная речь; 0,42243— художественная проза; 0,42509— драматургия. Как видим, и здесь явно выражена тенденция к понижению процента гласных в поэзии (а также иных функциональных стилях, ориентирующихся на письменную форму языка) и повышению в художественной прозе и драматургии, ориентирующихся на устную форму языка. Мы более осторожно, чем украинские авторы, формулируем наши выводы — понижение содержания гласных характерно, конечно, не для всякой

[6] В. И. Перебейнос. Частота и сочетаемость фонем современного украинского языка. Киев, 1964.

поэтической речи, а (на польском материале) для более традиционных стиховых форм.

Другим важным моментом, характеризующим поэзию раннего Тувима на фонологическом уровне, является повышение процента непрерывных согласных (в 17 стихотворениях частота непрерывных выше соответствующей частоты в прозе и в шести совпадает с ней, а лишь в двух меньше). Одно из этих двух стихотворений — «Podmuch wiosny» (процент непрерывных 13,8) построено на повышенном употреблении сонантов и плавных, создающем эффект певучести, эвфонии, передающий на звуковом уровне образ весны, радости. Соответственно в этом стихотворении резко снижено содержание класса свистящих и шипящих (до 0,061). Другое стихотворение — «Rzuciłbym to wszystko» (процент непрерывных 12,2) построено в темпе стаккато — соответственно большой удельный вес в звуковой структуре занимают смычные *ppbb́* и *td*, образующие аллитерирующую сетку.

Таким образом, увеличение содержания непрерывных согласных можно считать другой отличительной особенностью поэзии раннего Тувима, причем это увеличение настолько последовательно проводится, что именно снижение доли непрерывных воспринимается как отмеченное.

Среди непрерывных наиболее значительное повышение отличается для периферийных *ff vv́* и *x*. То, что в стихах раннего Тувима существенную роль играет именно признак «непрерывность», а не, положим, «яркость», доказывается тем, что процент аффрикат все время колеблется около среднего значения для прозы, в то время как именно этот класс согласных можно считать характерным для польского языка (средняя частота в польской прозе — 0,050 по нашим данным по сравнению с 0,024 в чешском и 0,022 в русском языках по данным Г. Кучеры). Поэтому небольшое увеличение процента непериферийных непрерывных *sśšzźž* становится значимым не в сочетании с аффрикатами, а в сочетании с другими непрерывными.

Среди гласных заметно уменьшение доли компактного *a* при сохранении соответствующих величин для некомпактных гласных.

Отмечая эти различия между частотными фонологическими структурами прозаических и поэтических текстов польского языка, мы не решаемся обобщать их для польской поэтической речи вообще. Однако, по-видимому, они могут быть приняты в качестве вероятной рабочей гипотезы при описании стиха раннего Тувима; особенно значимым представляется существенное увеличение частот непрерывных согласных в этих поэтических текстах.

В случайной выборке 25 стихотворений разных сборников нивелируются индивидуальные звуковые особенности каждого стихотворения и средние величины частот отражают лишь общее, присущее всем этим текстам.

С другой стороны, когда мы берем подряд все стихотворения одного цикла, объединенные не только тематически, но и в плане звуковой фактуры, то индивидуальные особенности каждого стихотворения оказываются повторяющимися в других, и средние значения частот показывают сильное отличие от частот, полученных в случайно подобранных текстах.

Именно так обстоит дело в двух рассматриваемых стихотворных циклах Тувима. Цикл из 7 стихотворений, предшествующий циклу «Słopiewnie», обнаруживает развитие и усиление тенденций, отмеченных для случайной выборки 25 стихотворений. Доля гласных — такая же, как и в случайной выборке (меньше, чем среднее значение для прозаических текстов польского языка), доля компактного *a* еще больше снижена, а доля гласных низкой тональности увеличена. Это увеличение процента периферийных гласных коррелирует с резким увеличением процента периферийных согласных (до 0,182), при этом периферийные любого способа образования обнаруживают в этом цикле тенденцию к увеличению частотности. С другой стороны, продолжится тенденция к возрастанию доли непрерывных согласных (до 0,197) независимо от места образования. В случаях наиболее резкого увеличения частоты класса фонем наблюдается значительная однородность этого увеличения по отдельным стихотворениям цикла; так, например, во всех семи стихотворениях частота класса непрерывных заведомо больше средней по прозе: в шести стихотворениях частота класса *ffvú*, в котором пересекаются существенные для данного цикла признаки периферийного места и непрерывного способа образования, больше средней по прозе, и лишь в одном стихотворении эти частоты совпадают; в пяти стихотворениях частота периферийных превышает среднюю по прозе, в одном совпадает с нею и в одном — меньше (речь идет о стихотворении «Świt», в котором наблюдается аллитерация свистящих и шипящих — в соответствии с названием стихотворения): в четырех стихотворениях частота класса прерывных периферийных *śśžźž* больше прозаической и в трех — совпадает с нею.

Таким образом, подтверждается интуитивное представление о том, что стихотворения, входящие в данный цикл, близки не только по тематике, но разделяют и сходные особенности звуковой фактуры, одинаковым образом отличаясь от прозаического уровня, а также от «общестихового» (для раннего Тувима) уровня. Именно в этом цикле поэт развивает «древнеславянские» мотивы языческого буйства природных стихий, что в звуковом плане представлено преобладанием шумных согласных (доля несогласных уменьшена по сравнению с общеязыковым уровнем).

В цикле «Słopiewnie» поэт решал несколько иную тематическую задачу — выявление в польской речи его первоначального, скрытого, древнеславянского пласта, восстановление истинного, певучего слова (Słо-piewnie!). Эта задача решается на разных язы-

ковых уровнях; на лексическом уровне употребляется большое количество слов, построенных по древнеславянскому образцу так, как Тувим его себе представлял. Этот образец сочетает в себе слова «диалектно-фольклорного» типа с окончанием на палатализованный согласный, сонант или плавный: ziel, jasnoziel, jarzeń, czerwoń, dziwierz ($rz < \acute{r}$), glaź, topiel и т. п. и слова, описывающие реалии, которые могли бы встретиться у древних славян — сюда входят такие слова, как dziewanna, dziewna, kniaziewna, księźawiec. При этом замечательно, что для передачи образа древности поэт использует слово kniaziewna (ср. русское *князь*), хотя по своему фонетическому облику оно ничуть не древнее польского księżna, сохранившего старый носовой. Эта ориентация на внепольские языковые образцы для передачи «древнеславянского духа» особенно проявляется в стихотворении «O mowie rosyjskiej», где передается уже на морфонологическом уровне и русское полногласие (żurczałowo, pieczałowo) и йотированное окончание прилагательного (tiewnaja). Модель с йотом между гласными в прилагательных и наречиях воссоздается и в слове dunajewo.

Стремясь пе́редать звонкость, певучесть, плавность «древнеславянской речи», поэт регулирует звуковую фактуру стиха: в противоположность рассмотренным выше стихам (как подобранным случайно, так и предшествующему циклу) здесь значительно повышается процент гласных (главным образом за счет *a*) и несогласных, среди которых количество носовых примерно соответствует среднему в прозе, а количество плавных резко увеличивается. Увеличение доли гласных и несогласных коррелирует здесь с резким повышением процента звонких согласных: если в случайной выборке 25 стихотворений доля согласных, произносимых без участия голоса (0,246), почти совпадает с аналогичными данными для цикла из семи стихотворений (0,245), то в «Słopiewnie» она падает до 0,180. Обращает на себя внимание значительное повышение процента палатализованных фонем в стихотворениях «Słopiewnie», что также выделяет этот цикл из всех рассмотренных текстов.

Все эти специфические отличия «Słopiewnie» от других поэтических текстов сближают весь этот цикл по статистическим характеристикам фонологического уровня с текстами русского языка — частота классов гласных плавных и палатализованных выше в русском языке, чем в польском. Таким образом, стихотворение Ю. Тувима «O mowie rosyjskiej», специально передающее фонологическими средствами впечатление от звучания русской речи, совпадает по статистическим характеристикам указанных трех классов фонем с остальными стихотворениями этого цикла (из шести стихотворений «Słopiewnie» в четырех частота гласных превышает среднепрозаическую, в одном совпадает с нею и в одном немного ниже ее; таково же распределение частот по стихотворе-

Таблица 3

Частоты классов фонем в поэтических текстах Ю. Тувима

Классы фонем	Средние по прозе	25 стихотворений раннего Тувима	7 стихотворений Epos — Rzeź brzóz	Słopiewnie
Гласные	0,414	0,403	0,404	0,429
o ō u	0,1275	0,127	0,139	0,115
e ē i	0,191	0,191	0,187	0,200
	0,0955	0,085	0,078	0,114
Несогласные	0,2125	0,221	0,202	0,231
m ḿ n ń	0,1055	0,114	0,092	0,103
r l ł̭ j	0,107	0,107	0,110	0,128
Согласные	0,3735	0,376	0,394	0,340
Периферийные	0,157	0,161	0,182	0,163
k ḱ g ǵ x	0,0595	0,062	0,067	0,038
k ḱ g ǵ	0,049	0,048	0,050	0,036
x	0,0105	0,014	0,017	0,002
p ṕ b b́ f f́ v v́	0,0975	0,099	0,115	0,125
p ṕ b b́	0,0485	0,046	0,049	0,043
f f́ v v́	0,049	0,053	0,066	0,082
Непериферийные	0,2165	0,215	0,212	0,177
t d c ʒ ć ʒ́ č ǯ	0,117	0,112	0,098	0,092
t d	0,067	0,062	0,051	0,047
c ć č ʒ ʒ́ ǯ	0,050	0,050	0,047	0,045
s ś š z ź ž	0,0995	0,103	0,114	0,085
Палатализованные	0,120	0,126	0,124	0,184
Звонкие	0,132	0,131	0,148	0,160
Яркие	0,1495	0,153	0,161	0,130
Непрерывные	0,159	0,170	0,197	1,179

ниям и для класса плавных; в пяти стихотворениях частота палатализованных существенно превышает соответствующую частоту в прозе — иногда более чем в два раза — и в одном совпадает с ней).

Цикл «Słopiewnie» отличается от предыдущих семи стихотворений тем, что в нем частота почти всех консонантных классов значительно ниже среднего уровня по прозе. Исключение составляет лишь класс *ffv́* и особенно звонкий представитель этого класса *v́*, последовательно пронизывающий все стихотворения «Słopiewnie». Во всех шести стихотворениях частота этого класса превышает прозаическую, причем иногда почти вдвое, что делает этот консонантный класс специфическим для всех разобранных

нами поэтических текстов и отличает их от текстов прозаических.

Проведенный разбор статистической структуры ряда стихотворений Ю. Тувима подтверждает, что в плане выражения поэтических текстов заключены значительные ресурсы. Эти ресурсы состоят, как нам представляется, в большей свободе использования элементов фонологического уровня в поэзии по сравнению с прозой при том, что самые общие вероятностные закономерности звуковых цепей в поэзии остаются теми же, что и в непоэтических текстах — сравнительное содержание гласных, несогласных и согласных в любом тексте одинаково: первое место занимают гласные, затем согласные, а затем несогласные.

Подобное исследование звуковой фактуры польского стиха показывает, что данные о частоте фонологических классов могут иметь и чисто практическое значение для смысловой интерпретации поэтических текстов. Исходя из априорной предпосылки, обратной той, с которой мы начали исследование,— т. е. о различии (а не о совпадении) частот фонологических классов в отдельных текстах, можно прийти к пониманию связи значения и звучания в стихе. Исследование границ такого различия (а тексты Ю. Тувима как раз и представляют такой граничный случай) должно, с другой стороны, помочь установить стабильность, уже не зависящую от факторов стиля и проч.

Отсюда также можно перейти к выявлению специфических фонетических характеристик отдельных текстов (что особенно важно при изучении поэтической и вообще художественной речи), групп текстов, произведений одного автора и т. п.

Описанный выше пример практического приложения сведений о частотности фонологических элементов не исчерпывает, разумеется, возможностей этих приложений. В частности, могут оказаться полезными подробные исследования фонологического заполнения частотной сетки (подобно тому, как у нас это проделано для польского языка). Эти данные могут быть использованы в самых различных целях — типология, дешифровка и проч.

§ 3. Некоторые предложения относительно интерпретации данных по статистике парной встречаемости фонем

Настоящая работа в своей основной части посвящена описанию и интерпретации статистики встречаемости единичных фонем и классов фонем. Рассматриваемая нами совокупность польских текстов представляла собой последовательность событий, каждое из которых состояло в однократной встречаемости отдельной фонемы (или класса фонем). В то же самое время значительный интерес для фонологии представляют данные о встречаемости нескольких фонем — пар, троек, четверок, пятерок и т. п. В этом

случае событием считается не встречаемость одной фонемы, а встречаемость двух (трех, четырех, пяти и т. д.) фонем подряд. Не приходится сомневаться в том, что подобные данные имеют огромное значение при описании дистрибуции фонем и могут быть использованы для построения статистических моделей синтагматических единиц на фонологическом уровне (слог, фонетическое слово). До сих пор все описания дистрибуции фонем строились по принципу «да — нет»: возможно ли данное сочетание фонем в данном языке или нет. Естественно, что подобная картина, при всей ее несомненной ценности на первом этапе исследования, является одномерной, упрощенной. Введение в подобные модели информаций о том, насколько часто или редко встречается данное сочетание фонем, значительно приближает ее к реальной языковой действительности.

Отметим, что при рассмотрении статистической дистрибуции на уровне пар фонем мы можем в какой-то степени опираться на априорное интуитивное представление о частости или редкости того или иного сочетания двух фонем: интуитивно мы чувствуем, что пара (*t a*) будет в польском языке встречаться чаще, чем пара (*f l*). Однако такое априорное постулирование возможно лишь для самых грубых случаев — пара «согласный + гласный (неносовой)» будет встречаться чаще, чем пара «согласный + согласный». В огромном большинстве случаев мы заранее ничего не сможем сказать о взаимной частости или редкости встретившихся пар фонем. Это связано прежде всего с тем, что при переходе к совместной встречаемости фонем мы резко увеличиваем число учетных единиц фонологического уровня. Достаточно сказать, что число разрешенных пар фонем для польского языка —700. Число разрешенных троек (разрешенных с чисто фонологической точки зрения) будет, по-видимому, еще бóльшим. .В результате число учетных единиц начинает приближаться к тому, что в словаре. Соответственно при рассмотрении совместной встречаемости фонем можно ожидать, что однородность будет меньшей, чем для единичных фонем. Отсюда вытекает, что привлечение конкретных числовых данных по встречаемости нескольких фонем в текстах при описании дистрибуции фонологических элементов еще менее обосновано статистически, чем указание конкретной частоты фонем при описании парадигматики.

В идеале следовало бы провести для сочетания нескольких фонем то же исследование, что мы провели для единичных фонем: определить элементы с более стабильной и менее стабильной частотой, рассмотреть частотную сетку на этом уровне, ее заполнение элементами и проч. Однако объем чисто технической работы в этом случае был бы столь грандиозен, что это было бы под силу лишь целому коллективу исследователей, вооруженному вычислительной техникой: при сравнении между собой различных выборок требовалось бы, чтобы каждая из них содержала все n-фонемные

сочетания, а поскольку число таких сочетаний очень велико, то объем выборок должен был бы быть фантастическим. Уже хотя бы поэтому надлежит искать других путей, не сопряженных с таким объемом работы.

В настоящем разделе мы предлагаем некоторый опыт интерпретации частот двухфонемных сочетаний. Эта интерпретация должна рассматриваться лишь как модель возможных дальнейших исследований, а фонологические выводы, полученные нами,— лишь как гипотезы для дальнейшей проверки. Наша интерпретация представляет собой попытку различения заранее выделенных фонологических непересекающихся классов (процедуру выделения см. в главе III) на основе некоторых достаточно грубых статистических показателей встречаемости этих классов совместно с другими классами. Иными словами, мы исследуем, как отражается в статистическом плане априори известное отличие выделенных классов друг от друга. В качестве основы интерпретации нами была избрана, с одной стороны, модель Ципфа, постулирующая превышение частоты класса, идентифицируемого отсутствием данного признака над частотою класса, идентифицируемого его присутствием, а с другой стороны, модель Харрари и Пэйпера [7], позволяющая вводить определенные количественные параметры при описании дистрибуции фонем.

При исследовании встречаемости двухфонемных сочетаний мы не проводили проверки полученных частот на однородность. Это было связано, во-первых, с тем, что принципиальная неоднородность текстов относительно частот парных сочетаний фонем была ясна с самого начала, а во-вторых, с крайней ограниченностью и случайностью обследованного материала. Поскольку подсчет двухфонемных сочетаний был невозможен без помощи вычислительной техники, а в нашем распоряжении она была лишь короткое время, объем обследованного материала весьма невелик: подсчеты велись по 1 главе повести Я. Ивашкевича «Девушка и голуби», привлекавшейся одновременно и для подсчета встречаемости единичных фонем, а также по статье А. Шаффа «Marksism a filozofia człowieka», опубликованной в 1961 г. в газете «Przegląd kulturalny».

Общий объем обследованного материала, включая пробел между фонетическими словами,—30 903 знака, без учета пробела — 27 192 знака. Соответственно общее число обследованных пар элементов (включая пробел) равно n — 1 или 30 902. В этом подсчете мы учитывали пробел между словами, так как сочетания пробела с последующей и предшествующей фонемой дают сведения о распределении фонем в начале и конце слова. Мы произвели укрупнение единиц (аналогично тому, как вместо единич-

---

[7] F. Harrary and H. Paper. Toward a general calculus of phonemic distribution. «Language», 1957, v. 33, № 2.

ных фонем мы стали рассматривать классы фонем) и вместо сочетания 42 элементов друг с другом мы будем рассматривать совместную встречаемость следующих классов: пробел, гласные, сонанты, плавные, *kǩ gǧ, x, pṕ bḃ, fḟ vv́, td, sśš zźž, cćč ʒʒ́ǯ*. Сочетания из 11 элементов по два дают нам 55 учетных единиц, т. е. количество, соизмеримое с количеством фонем в языке.

Частотность этих 55 пар классов мы не будем описывать в виде распределения в порядке убывания частоты, как мы это делали для фонем. Для интерпретации частот совместной встречаемости классов фонем мы применим другую процедуру. Еще раз подчеркнем, что полученные результаты (в отличие от результатов по встречаемости единичных фонем и классов фонем) никоим образом не могут претендовать на окончательность.

Мы будем исследовать не относительную частоту совместной встречаемости классов фонем и распределение этой частоты, а более грубое отношение «больше — меньше», т. е. констатируется, что сочетание АВ встретилось чаще (или реже, или столько же раз), чем сочетание АС, а затем это превышение (или равенство) интерпретируется. Между собою сравниваются сочетания, имеющие один общий элемент (АВ — АС или АВ — СВ), что позволяет объединять данные о встречаемости перед данным классом или после него.

Каждый класс характеризуется набором различительных признаков, поэтому сведение результатов по фонемам в более крупные единицы дает своего рода синтаксис различительных признаков в пределах пары фонем.

Цифры относительной частоты совместной встречаемости классов фонем не рассматриваются: это представило бы результаты в неуместно точной форме, в то время как на самом деле более или менее точным в нашем подсчете является не процент данного сочетания в тексте, а его ранг, порядок по убывающей частоте. Цифры относительной частоты имеют смысл только тогда, когда они отражают положение в более общей совокупности (именно поэтому частота и относительна — мы вычисляем наши данные по выборке, а потом относим их ко всей совокупности). Здесь же относительные частоты отражали бы лишь два исследованных текста, поэтому не имело смысла стремиться к ложной точности.

Например, в начале слова, по нашим результатам, гласные встречаются 703 раза, а в конце слова — 2348 раз. Не высказывая мнения о том, насколько точно (в статистическом смысле) такое превышение, мы все же не можем не отметить его величину, не можем не констатировать (опять-таки, не уточняя этого высказывания), что в начале слова гласные встречаются реже, чем в конце. Это высказывание, сделанное в самой общей форме, и будет гипотезой, подлежащей проверке в дальнейших исследованиях.

Таблица 4

Встречаемость классов фонем польского языка в терминальных позициях

Классы фонем	Начало слова	Конец слова	Классы фонем	Начало слова	Конец слова
Гласные	703	2348	Периферийные некомпактные	809	69
Негласные	3007	1362			
Несогласные	830	699	Периферийные некомпактные непрерывные	*f f́ v v́* 381	*f* 21
Носовые несогласные	*m ḿ n ń* 447	*m n ń* 349			
Плавные несогласные	*r l ụ j* 383	*r l ụ j* 350	Периферийные некомпактные прерывные	*p ṕ b b́* 428	*p* 48
Согласные	2177	663			
Периферийные согласные	1056	295	Непериферийные согласные	1121	368
Периферийные компактные	247	226	Непериферийные непрерывные	*s ś š z ź ž* 627	*s ś š* 148
			Непериферийные прерывные	494	220
Периферийные компактные непрерывные	*x* 39	*x* 129	Непериферийные прерывные яркие	*c ć č ʒ ʒ́ ǯ* 142	*c ć č* 144
Периферийные компактные прерывные	*k ḱ g ǵ* 208	*k* 97	Непериферийные прерывные тусклые	*t d* 352	*t* 76
			Всего	3710	3710

Подобным образом мы постараемся описать распределение классов фонем в польском языке.

Начнем с описания встречаемости классов фонем совместно с пробелом, т. е. в начале и конце фонетического слова.

Первой бросающейся в глаза особенностью распределения частот классов польских фонем в зависимости от начала или конца слова является только что упоминавшееся преобладание гласных в конце слова и негласных — в начале. «Вокалический» характер конца слова в тексте подчеркивается еще и тем, что в этой позиции (в отличие от начальной) число плавных и носовых сонантов превышает число согласных, хотя абсолютное число «несогласных» меньше в конце, чем в начале.

Сравнение абсолютных цифр встречаемости каждого класса в обеих позициях дает в общем немного: в начальной позиции все классы негласных фонем (за исключением двух случаев, о которых речь ниже) встречаются чаще, чем в конечной. Поэтому такое сравнение частот каждого класса друг с другом должно быть дополнено сравнением всего распределения частот в обеих позициях.

Для этого выберем точку отсчета — такое распределение, которое было бы «нормальным», по сравнению с которым любое

отклонение будет значимым. В качестве такой модели мы избрали модель Дж. К. Ципфа, согласно которой класс фонем, идентифицируемый отрицательным значением различительного признака, встречается чаще, чем класс фонем, идентифицируемый положительным значением признака. Будем считать эту модель исходной и проверим, насколько она выполняется в терминальных условиях. Для начальной позиции это правило выполняется в пределах класса периферийных согласных:

р (непериферийные) > р (периферийные)
р (некомпактные непериф.) > р (компактные периф.)
р (*kǩgǵ*) (прерывные) > р (*x*) (непрерывный)
р (*pṕbḃ*) (прерывные) > р (*ff vv́*) (непрерывные)
р (*pṕbḃ*) (некомпактные > р (*kǩgǵ*) (компактные)
р (*ff vv́*) (некомпактные) > р (*x*) (компактный)

Для конечной позиции общее соотношение непериферийных и периферийных остается тем же: непериферийных больше, чем периферийных. Однако внутри класса периферийных согласных соотношения меняются:

р (*pf*) (некомпактные) < р (*kx*) (компактные)
р (*k*) (прерывный) < р (*x*) (непрерывный)
р (*p*) (некомпактный) < р (*k*) (компактный)
р (*f*) (некомпактный) < р (*x*) (компактный)

Отношение частот периферийного некомпактного прерывного *p* к соответствующему непрерывному *f* то же, что и в начале слова: р (*p*) > р (*f*).

Тот факт, что заднеязычные согласные встречаются в конце слова чаще, чем губные, выделяет конечную позицию как специфическую для заднеязычных согласных. Здесь мы обнаруживаем первый дистрибутивный критерий для разделения классов согласных. Конечная позиция особенно специфична для класса, представленного одной фонемой *x*. Эта фонема является исключением из общей картины встречаемости согласных в терминальных позициях: абсолютная частота *x* в конце слова выше, чем в начале, а также выше, чем у *k* (т. е. фонемы, отождествляемой отрицательным значением признака «непрерывность») в конце слова. Эти данные выделяют фонему *x* как специфическую фонему именно конечной позиции.

Отметим, что резкое уменьшение частотности согласных в конце слова может быть связано, в частности, с оглушением звонких согласных и с отсутствием на конце слова большинства палатализованных коррелятов (кроме *ńść*) в отличие, например, от русского или словацкого языков. Фонема *x* — единственная среди всех согласных фонем, стоящая вне противопоставлений по палатализованности — непалатализованности и звонкости — глухости, поэтому конечная

позиция не является для нее ограничительной. Это может объяснить повышенную встречаемость *x* именно в конечной позиции.

Разумеется, чисто фонологическое объяснение различий во встречаемости классов фонем в терминальных позициях недостаточно. Повышенная встречаемость гласных именно в конце слова связана, в частности, и с морфологическими причинами: много морфологических формантов как в имени (например, окончание именительного падежа множественного числа), так и в глаголе (также окончание множественного числа) — это гласные. Аналогичным образом дело обстоит и с *x*, которое чаще всего встречается не в корневых морфемах, а в окончании именных флексий множественного числа *-ach*, *-ich*. С другой стороны, отметим, что губные фонемы — это специфические фонемы начальной позиции — абсолютная частота губных в начале слова превышает частоту этих фонем в конце более чем в десять раз (заднеязычные в начале и конце разнятся на гораздо меньшую величину: 247 против 226).

Подобная локализация позиции в слове за губными и *x* любопытным образом соответствует способу произношения: в начале слова (среди периферийных) преобладают фонемы, артикулируемые в начале речевого аппарата, а в конце — фонемы, артикулируемые в глубине.

Если правило Ципфа выполнялось для периферийных согласных в начальной позиции, то в отношении непериферийных согласных начальная и конечная позиции меняются местами:

в начальной позиции		в конечной позиции	
p (*tdczćźčž*) <	p (*sššzźž*)	p (*tcćč*) <	p (*sšš*)
прерывные	непрерывные	прерывные	непрерывные

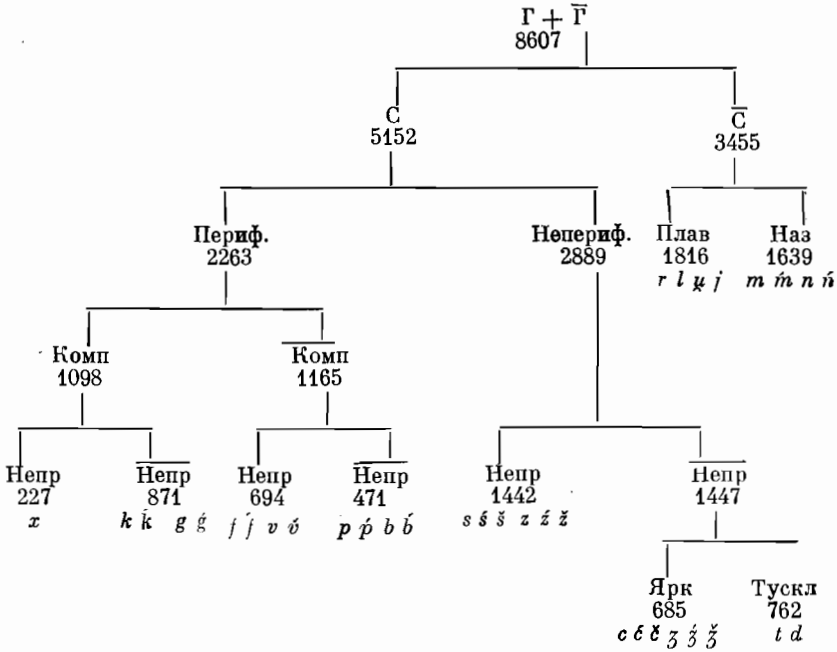
Для тусклых и ярких непериферийных имеет место обратное:

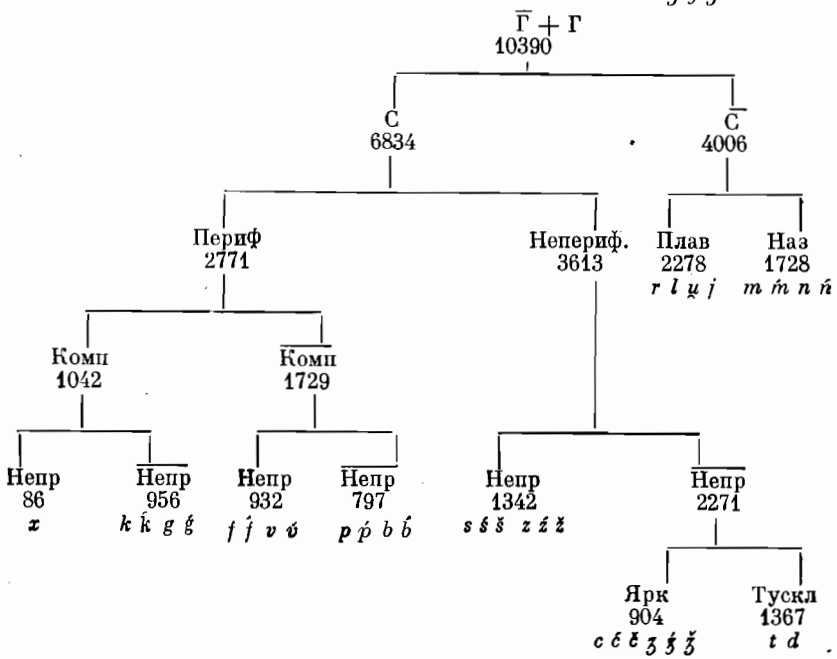
в начальной позиции		в конечной позиции	
p (*td*) >	p (*cćčźžǯ*)	p (*t*) <	p (*cćč*)

Таким образом, в начальной позиции среди непериферийных согласных выделяется класс *sššzźž*, отступающий от модели Ципфа, т. е. встречающийся значительно чаще прерывных *tdcćč źžǯ*, а в конечной позиции — класс *cćč*, также отступающий от этого правила и встречающийся чаще тусклых *td*. Частота *cćč* в конечной позиции к тому же не уменьшена по сравнению с начальной позицией, что выделяет этот класс с точки зрения распределения. Класс непериферийных непрерывных *sššzźž* может рассматриваться как специфический для начальной позиции, а класс *cćč* — как специфический для конечной позиции.

Итак, сравнение частот встречаемости классов фонем в начале и конце слова позволяет сделать следующие предварительные выводы, которые могут выступить в качестве рабочих гипотез, подлежащих проверке при последующих исследованиях:

Встречаемость пар «гласный — согласный» и «согласный — гласный» в польском языке





— в начале фонетического слова чаще встречаются согласные, а в конце гласные;
— в начальной позиции частотность классов периферийных согласных фонем подчиняется правилу Ципфа; в конце слова это правило нарушается;
— фонема *x* чаще встречается в конце слова, чем в начале;
— губные фонемы чаще встречаются в начале слова, чем в конце;
— в начальной позиции правило Ципфа нарушается для соотношения непериферийных непрерывных и непериферийных прерывных;
— в конечной позиции это правило нарушается для соотношения ярких и тусклых непериферийных;
— аффрикаты чаще встречаются в конце слова, чем в начале.

Нарушения правила Ципфа и соотношения частот в начальной и конечных позициях позволяют статистически разграничить различные классы согласных фонем в начале и конце слова.

Так же, как и в терминальных позициях, в данном случае негласные встречаются чаще, когда они являются первым элементом пары (в том случае $\overline{\Gamma}\#$, здесь $\overline{\Gamma}\Gamma$).

Правда, в случае совместной встречаемости гласных и согласных гласные также встречаются чаще в типе $\overline{\Gamma}\Gamma$. Так же, как и в терминальных позициях, в рассматриваемом случае общее правило таково, что частотность классов согласных там, где они предшествуют гласным, больше чем там, где они следуют после них, так что в некотором смысле гласная может считаться аналогичной паузе. Исключение составляет тот же периферийный компактный непрерывный *x*, который встречается чаще после гласных, чем перед ними. Соответственно общая частотность класса компактных периферийных также больше после гласных, чем перед ними (правда, это превышение незначительно).

Таким образом, согласный *x* может быть на основании наших данных охарактеризован как статистически преобладающий в исходе. Этот тип согласных как бы закрывает слово или фонемную пару, встречаясь гораздо чаще в конце, чем в начале.

Так же, как и *x*, класс непрерывных непериферийных согласных *sšžzź* встречается после гласных чаще, чем перед ними. Как будет показано ниже, это связано с особой ролью этого класса в консонантных сочетаниях. В отличие от *x*, частотность которого статистически ограничена позициями исхода, класс *sšžzź* не проявил таких особенностей в терминальной позиции.

Еще одна особенность встречаемости гласных и согласных — это то, что в обоих типах — $\Gamma\overline{\Gamma}$ и $\overline{\Gamma}\Gamma$ — периферийные некомпактные непрерывные *ffvv́* встречаются чаще, чем соответствующие прерывные *pṕbb́*, что нарушает гипотезу Ципфа о зависимости частоты от наличия — отсутствия признака. Для остальных консонантных классов эта гипотеза на уровне сочетаний гласных и согласных подверждается.

Таким образом, сравнение частот классов фонем по двум критериям (совпадение с гипотезой Ципфа и совпадение с общей тенденцией распределения) выделяет из всех консонантных классов классы непрерывных фонем. Каждый из этих классов обнаруживает отклонения в частоте по одному из двух критериев, что, в свою очередь, выделяет непрерывные согласные из всех других фонем на уровне статистической дистрибуции.

Применяемый здесь способ сравнения частот зеркальных пар элементов по критерию совпадения и общей тенденцией распределения «p (α_i β_i), как правило, больше p (β_i α_i)» можно переформулировать следующим образом. Если частота $\alpha\beta$ существенно больше частоты $\beta\alpha$ (к сожалению, здесь приходится ограничиваться лишь интуитивными оценками такого превышения), это значит, что α с большей вероятностью предсказывает β, чем β предсказывает α. Это свойство можно назвать «предсказуемостью» α в отношении β. Соответственно фонема *x* обладает свойством предсказуемости в отношении конца слова; гласные обладают большей предсказуемостью по отношению к *x*, чем наоборот; класс *kǩgǵ* (как и многие другие классы) обладает предсказуемостью в отношении гласных и т. п. С другой стороны, гласные обладают большей предсказуемостью в отношении класса *sśšzźž*, чем этот класс в отношении гласных.

Легко заметить, что свойство сравнительной предсказуемости двух классов (элементов) относительно друг друга отчасти сравнимо с логическим свойством симметричности.

О логическом свойстве симметричности двух элементов α и β говорят в том случае, когда при истинности $\alpha\beta$ истинно и $\beta\alpha$. Соответственно две фонемы связаны свойством симметричности, когда возможно сочетание AB и сочетание BA. Введение количественной стороны в характеристику сочетаемости уточняет выделяемые логические отношения фонем (или классов фонем). Соответственно можно говорить о статистической симметричности (когда p/AB/ = = p/BA/) и о статистической предсказуемости A в отношении B, когда при возможных AB и BA p/AB/ $>$ p/BA/.

Свойство предсказуемости позволяет углубить классификацию множества пар классов фонем.

Упорядочим теперь множество гласных, встречающихся после каждого класса согласных и перед ним, в порядке убывающей частоты и посмотрим, в отношении каких гласных консонантные классы обладают большей предсказуемостью (табл. 5).

Поскольку в целом невокальные классы обладают большей предсказуемостью в отношении гласных, чем гласные в отношении негласных, а также поскольку в начале слов преобладают негласные, и, следовательно, началом звуковой цепи в польском языке статистически можно считать CV, сделаем предварительное допущение, что в паре $\Gamma\bar{\Gamma} \vee \bar{\Gamma}\Gamma$ различительные признаки негласного больше связаны с различительными признаками после-

Таблица 5

Порядок встречаемости гласных фонем перед данным классом	Класс негласных фонем	Порядок встречаемости гласных фонем после данного класса	Порядок встречаемости гласных фонем перед данным классом	Класс негласных фонем	Порядок встречаемости гласных фонем после данного класса
e i a o u	*m m̂ n ń*	*a e i o u õ ẽ*	*o a i u e*	*f f́ v v́*	*i a e o õ u ẽ*
a e o i u	*r l ų j*	*a e o i u õ*	*o i e a u õ ẽ*	*s ś š z ź ž*	*i e a o u õ ẽ*
e a o i u	*k k̂ g ĝ*	*o a e i u õ ẽ*	*a e o i u*	*t d*	*e o a i u õ ẽ*
i a e o u õ ẽ	*x*	*o a u õ i e*	*e i a o u*	*c ć č ʒ ʒ́ ǯ*	*e i a o u õ ẽ*
o e i a u	*p ṕ b b́*	*o e i a u õ ẽ*			

Примечание. Гласные упорядочены по убывающей частоте слева направо.

дующего, чем предыдущего гласного. Исходя из этого допущения, сгруппируем консонантные классы согласно рангу гласных фонем после них. На первом месте в строке стоит самая частая гласная, встречающаяся после данного класса, имеющая первый ранг, второй ранг имеет следующая по частоте гласная и т. п.

По последующей гласной первого ранга объединяются классы: *m m̂ n ń* и *r l ų j* (по *a*), *k k̂ g ĝ*, *x* и *p ṕ b b́* (по *o*), *f f́ v v́* и *s ś š z ź ž* (по *i*) *t d* и *c ć č ʒ ʒ́ ǯ* (по *e*). Мы видим, что группировка вполне естественная — объединяются классы, идентифицируемые общими различительными признаками. Равным образом гласные первого ранга обладают существенными признаками, объединяющими их с предшествующими им невокальными классами.

Фонема *i* характеризуется в терминах артикуляционных признаков высоким подъемом и передним рядом, т. е. она локализуется в передней части речевого аппарата при сравнительно малой величине резонирующей полости, которая соответственно смещена вперед.

Подобные же признаки характеризуют непрерывные классы *f f́ v v́*, *s ś š z ź ž*. Класс *f f́ v v́* локализуется в самом начале речевого аппарата. Поскольку фонемы, образующие этот класс, непрерывные, органы речи находятся в напряжении, что приводит к уменьшению резонатора. Класс свистящих и шипящих артикулируется в передней части полости рта и также характеризуется высоким подъемом языка.

Другим характерным примером совместной встречаемости гласных и согласных является сочетание классов *k k̂ g ĝ*, *x* и *p ṕ b b́* с последующей гласной первого ранга *o*. Члены консонантных классов локализуются в крайних противоположных точках полости рта: сзади и спереди. При этом объем резонатора велик. В терминах акустических признаков эти согласные фонемы характеризуются

признаком «grave» («тупой» или «периферийный»). То же самое следует сказать и о фонеме *o*, которая локализуется в задней части полости рта и произносится при низком положении языка, т. е. с большим резонатором.

Классы непериферийных прерывных *td* и *cćčzźž* имеют последующую гласную первого ранга — *e*.

Согласные фонемы характеризуются по акустической терминологии признаком «acute» («острый», «непериферийный»), равно как и гласная *e*. Это свидетельствует о том, что и согласные *td cćčzźž* и гласная *e* локализуются не в самом начале ротовой полости, а чуть дальше к середине. Прерывность согласных связывается с движением языка вниз (по сравнению с *sśšzźž*) и соответственно с увеличением раствора. Именно эти признаки характеризуют фонему *e* по сравнению с *i*.

Неконсонантные классы *mṁnń* и *rlu̯j* характеризуются в разных фонологических системах по-разному, но везде их объединяет то, что они противопоставляются как гласным, так и согласным. Гласная фонема *a* в польском литературном языке противопоставлена всем остальным гласным как компактный — некомпактным и с трудом поддается идентификации в терминах бинарных различительных признаков. Таким образом, предпочтительная сочетаемость несогласных с последущим *a* также основана на выделении некоторых общих признаков у гласных и согласных.

На основании вышеприведенного разбора кажется весьма вероятным, что статистическая сочетаемость консонантных классов с последующими гласными идет по линии подбора в первую очередь гласных, содержащих общий признак с согласным, при этом релевантен признак «grave — acute». «Тупые» согласные в первую очередь сочетаются с «тупыми» гласными, а «острые» согласные — с «острыми» гласными. Этой общей тенденции на первый взгляд противоречит положение класса *ff vó*, который обычно идентифицируется как «периферийный» («тупой») и после которого, по нашим данным, статистически преобладают «острые» гласные. Однако, по-видимому, подобное преобладание можно рассматривать как указание на то, что класс *ff vó* относится к «острым» согласным скорее, чем к «тупым». Артикуляторные причины этого мы привели выше. Действительно, поскольку основным признаком «тупых» является форма резонирующей полости, близкая к резонатору Гельмгольца, *ff vó* могут, вследствие напряженности речевых органов (и связанного с этим уменьшения резонатора), быть отнесены к «острым» согласным.

Если рассмотреть сочетаемость консонантных классов уже не с одной последующей гласной первого ранга, а с двумя самыми частыми гласными, то картина в основном останется та же.

Классы *mṁnń* и *rlu̯j* объединяются и по двум самым частым гласным *ae*. Классы *kḱgǵ* и *x* также объединяются по гласным двух первых рангов *o* и *a*. При этом ни *o* ни *a* не являются гласными

«непериферийными» — «острыми». В случае фонемы *x* принцип сочетаемости согласных периферийных («тупых») фонем с периферийными гласными проявляется еще более ярко: «острые» *i* и *e* занимают последние места в списке гласных, следующих после *x*.

Столь же характерна и сочетаемость классов «острых» фонем *sśšzźž* и *cćčzźž* с последующими гласными. Эти два класса объединяются тем, что после них на первых двух местах стоят именно «острые» гласные *e*, *i*. Такое статистическое распределение объединяется еще и тем, что оба эти класса включают много палатальных фонем. Классы *pṕbb́* и *t d* объединяются по двум фонемам *e* и *o*, занимающим два первых места после этих классов. В обоих случаях согласные — смычные, их артикуляция не сопровождается напряжением органов речи, тело языка занимает низкое положение; это же характерно и для гласных.

Таким образом, признаки, характеризующие классы согласных, предсказывают соответствующие признаки у последующих гласных: признак «grave» у согласных статистически имплицирует признак «grave» у гласных, признак «acute» у согласных статистически имплицирует признак «acute» у гласных, а «несогласность + + негласность» имплицирует гласный *a*. Статистическая предсказуемость согласных в отношении гласных (p / AB / > p / BA /) коррелирует с фонологической предсказуемостью.

Рассмотрим теперь вокалическое окружение консонантных классов с обеих сторон. В целом набор гласных фонем, встречающихся перед согласными, меньше, чем после них. Есть, однако, два исключения, связанные с тем, что выше говорилось о количественном распределении типов Г $\overline{\text{Г}}$ и $\overline{\text{Г}}$ Г: перед *x* встречаются все гласные фонемы, а после него не встречается *ē*. Второе исключение состоит в том, что перед и после класса *sśšzźž* встречаются все гласные.

Если оставить в стороне тот факт, что набор гласных до и после консонантного класса неодинаков, и рассмотреть лишь порядок по убывающей частоте в обоих вокалических полях, то следующие консонантные классы обладают по отношению к гласным свойством, которое предлагается назвать ранговой симметричностью (ранговая симметричность — это одинаковый порядок элементов по убывающей частоте в α- и β-полях[8]: *rlu̯j*,*pṕbb́* и *cćčzźž*).

Рангово-симметричные консонантные классы не объединяются друг с другом общими признаками. Более того, объединение консонантных классов по гласным первого ранга в α-поле дает очень мало: по *e* объединяются классы *mm̀nń*, *kkgǵ* и *cćčzźž*, по *a* — классы *rlu̯j* и *t d*, по *o* — классы *pṕbb́*, *ff́ vv́*, *sśšzźž* и остается один класс (опять *x*), следующий чаще всего после *i*. Мы видим здесь, что признаки предшествующих гласных не предсказывают (или предсказывают весьма условно: после *o* чаще всего идут *ff́ vv́* и *pṕbb́* — губ-

[8] О понятиях α- и β-поле см. в указанной работе Харрари и Пэйлера.

ные, но также и *sśšzźž*) признаки последующих классов согласных так последовательно, как это имело место в случаях следования гласных после согласных.

Таким образом, в общем подтверждается предварительное допущение о том, что признаки классов согласных чаще предсказывают признаки гласных, чем наоборот.

Выделение рангово-симметричных по отношению к гласным консонантных классов имеет тот смысл, что на фоне этих классов резко выделяются рангово-несимметричные классы *x*, *ff vv́* и *sśšzźž*, т. е. классы непрерывных фонем, характеризующихся тем, что для этих классов набор фонем первых трех рангов в β-поле отличается от соответствующего набора в α-поле.

Для рангово-симметричных классов можно было считать справедливым, что различительный признак, статистически преобладающий в сочетании консонантного класса с последующей гласной, преобладает и в сочетании этого класса с предшествующей гласной; для классов *x*, *ff vv́* и *sśšzźž* верно обратное: фонема первого ранга в α-поле этих классов (т. е. в поле гласных, предшествующих данному классу) идентифицируется противоположным значением признака по сравнению с фонемой соответствующего ранга в β-поле.

Фонема 1-го ранга α-поля класса *x* — «острое» *i* при «тупом» *o* в β-поле; фонема 1-го ранга α-поля класса *ff vv́* — «тупое» *o* при «остром» *i* в β-поле, аналогично обстоит дело и в классе *sśšzźž*.

При этом характерно, что для классов *x* и *sśšzźž* частота предшествующих гласных больше, чем частота последующих гласных. Если применять свойство статистической предсказуемости, то окажется, что в тех случаях, когда гласные более предсказывают консонантный класс, чем этот класс предсказывает гласные, признак гласного («тупой — острый») предсказывает противоположное значение этого признака у согласных.

Таким образом оказалось, что изучение рангового порядка предшествующих гласных выделяет непрерывные фонемы как обладающие тем свойством, что перед ними чаще всего бывают гласные, идентифицируемые противоположным значением признака «тупой — острый», чем гласные, которые чаще всего следуют после этих классов.

Сочетаемость негласных с гласными позволяет отделить «тупые» (периферийные) согласные от «острых» (непериферийных) и непрерывные от прерывных. Те же признаки существенны и для сочетаемости негласных друг с другом.

Нами было отмечено 240 случаев совместной встречаемости плавных и сонантов друг с другом. Эти случаи распределяются следующим образом:

плавные	+	плавные	— 32
носовые	+	носовые	— 34
плавные	+	носовые	—162
носовые	+	плавные	— 12

Таблица 6

1-е место	Консонантная группа	2-е место	1-е место	Консонантная группа	2-е место
Носовые		Носовые	Плавные		Плавные
4	x	23	0	x	15
55	$k\,\acute{k}\,g\,\acute{g}$	27	47	$k\,\acute{k}\,g\,\acute{g}$	102
10	$f\,\acute{f}\,v\,\acute{v}$	61	9	$f\,\acute{f}\,v\,\acute{v}$	77
35	$p\,\acute{p}\,b\,\acute{b}$	22	41	$p\,\acute{p}\,b\,\acute{b}$	190
92	$s\,\acute{s}\,\check{s}\,z\,\acute{z}\,\check{z}$	191	98	$s\,\acute{s}\,\check{s}\,z\,\acute{z}\,\check{z}$	115
127	$t\,d$	113	57	$t\,d$	132
162	$c\,\acute{c}\,\check{c}\,\acute{z}\,\check{z}$	102	56	$c\,\acute{c}\,\check{c}\,\acute{z}\,\check{z}$	165

Сочетания несогласных с согласными представлены в табл. 6. Левая сторона таблицы показывает совместную встречаемость носовых сонантов и согласных. По признаку сравнительной предсказуемости непрерывные согласные сразу же отделяются от прерывных: непрерывные согласные x, $f\acute{f}v\acute{v}$ и $s\acute{s}\check{s}z\acute{z}\check{z}$ обладают большей предсказуемостью в отношении последующих сонантов, чем, наоборот, сонанты — по отношению к непрырывным согласным. В подобного рода сочетаниях сонанты функционируют аналогично прерывным согласным (ниже мы увидим, что сочетания типа «непрерывный + прерывный» — наиболее частотные среди согласных). Соответственно в случаях, когда консонантный класс является первым элементом пары, гипотеза Ципфа о том, что класс, идентифицируемый отсутствием признака, встречается чаще, чем класс, идентифицируемый положительным значением, нарушается. Правда, x + носовые встречается реже, чем $k\acute{k}g\acute{g}$ + носовые (23 против 27), но, если учитывать, что x представлено одним элементом, а $k\acute{k}g\acute{g}$ — двумя (kg), то превышение $k\acute{k}g\acute{g}$ над x окажется мнимым.

Гипотеза Ципфа выполняется в тех случаях, когда консонантный класс — второй элемент пары (прерывные встречаются чаще непрерывных).

Довольно четко отделяются друг от друга периферийные и непериферийные, причем в этом случае гипотеза Ципфа выполняется: непериферийные в качестве как первого, так и второго элемента пары встречаются примерно в полтора—два раза чаще периферийных (max. периферийных 61 — min. непериферийных 92).

Почти во всех случаях в сочетаниях носовых согласных с сонантами элементы $n\acute{n}$ (непериферийные) выступают гораздо чаще, чем $m\acute{m}$ (периферийные). Исключение составляет лишь сочетание элементов $m\acute{m}$ с классом $p\acute{p}b\acute{b}$, которое встречается чаще,

чем сочетание *nń* с *ṕpbb́* (27 против 8) — здесь имеет место естественная взаимная аттракция губных. Поскольку перед классом губных *ff vv́ mḿ* встречается крайне редко, можно сделать вывод, что признак смычности (прерывности) класса губных *ṕpbb́* имплицирует соответствующее сочетание признаков губности и смычности у сонантов. Таким образом, обнаруживается дистрибутивный критерий для отделения класса *ṕpbb́* от класса *kḱgǵ* (соответствующие прерывные непериферийные были выделены выше).

Ниже мы приводим таблицу, иллюстрирующую соотношение *mḿ* и *nń* в сочетаниях с согласными (см. табл. 7).

Таблица 7

Сонанты — 1-й элемент			Сонанты — 2-й элемент		
2	1		1	2	
	mḿ	*nń*		*mḿ*	*nń*
x	1	3	*x*	5	18
k ḱ g ǵ	4	51	*k ḱ g ǵ*	6	21
f f́ v v́	3	7	*f f́ v v́*	1	60
p ṕ b b́	27	8	*p ṕ b b́*	1	21
s ś š z ź ž	5	87	*s ś š z ź ž*	71	120
t d	3	124	*t d*	20	93
c ć č ʒ ʒ́ ǯ	21	141	*c ć č ʒ ʒ́ ǯ*	3	99

В большинстве случаев превышение встречаемости *n ń* над *ḿ m* очень большое. Исключение (помимо случая *ṕpbb́*) составляет встречаемость класса *śśžźž* с *mḿ*: 77 раз, что вполне соизмеримо с встречаемостью *śśžźž* плюс *nń* — 120. Здесь мы впервые встречаемся с явлением, которое весьма характерно для класса неперифериийных непрерывных *śśžźž* — встречаемость этого класса с прерывными (а *mḿ* должны быть отнесены к прерывным) значительно превышает соответствующие величины для других консонантных классов. Это может, в свою очередь, служить дополнительным дистрибутивным критерием (помимо уже названных) для выделения класса *śśžźž*.

Встречаемость консонантных классов с плавными характеризуется большей регулярностью, а следовательно, и большей трудностью разграничения разных консонантных классов. Согласные регулярно обладают большей предсказуемостью в отношении плавных, чем плавные в отношении согласных. Эта черта сближает плавные с гласными, где наблюдалось в общем то же самое.

Сравнение частот сочетаний «консонантный класс + носовой» и «консонантный класс + плавный» показывает, что в большинстве случаев частоты второго ряда превышают частоты первого ряда. Исключениями являются класс *śśžźž* и фонема *x*.

Гипотеза Ципфа выполняется для всех случаев в паре «плавный + консонантный класс», кроме случая R + *sśšzźž* (98) — R + *td* (условное обозначение плавных — R) (57) и для всех случаев в паре «консонантный класс + плавный» кроме случая *td* + + R (132) — *cćč zźž* + R (165). Обращает на себя внимание и то, что в парах с консонантным классом в качестве первого элемента наиболее частым является класс *pṕbb́* (190) при том, что в аналогичных парах с носовыми этот класс занимал последнее место по частоте (22). Таким образом, сравнение встречаемости согласных с носовыми и плавными дает ряд статистических дистрибутивных критериев, позволяющих разделять разные классы. Наибольшее количество отличий от других классов показывают классы *sśšzźž* и *x*. Среди остальных классов каждый по какому-то критерию можно отделить от других. Пожалуй, только класс *kḱgǵ* проявляет меньше всего особенностей, однако поскольку он выступает в паре с *x*, его также можно классифицировать.

В табл. 8 показано консонантное окружение классов согласных с обеих сторон. По числу классов согласных в правом и левом окружении (α- и β-поля) выделяются 7 рангов, расположенных в порядке убывания частоты соответствующих классов согласных, расположенных после данного класса или перед ним.

Следует отметить, что в таблице заполнено большинство мест (здесь мы не показываем традиционно описываемых явлений уподобления согласных, поскольку они достаточно известны, но, разумеется, любая графа, где совместно встречаются два консонантных класса, взаимносодержащие, например, глухие и звонкие согласные, должна читаться таким образом, что перед звонкими встречаются звонкие, а перед глухими — глухие и т. п.). Когда обычно составляются таблицы совместной встречаемости фонем (особенно согласных), то всегда остается вероятность того, что некоторые сочетания фонем не встретились, поэтому такие таблицы имеют тенденцию различаться от автора к автору.

В рассматриваемом случае укрупнение единиц исключило возможность того, что какое-либо сочетание консонантных классов осталось неучтенным. Невстретившиеся сочетания классов действительно невозможны в пределах фонетического слова, выделенного нами; это сочетания:

x + *x*		*x* + *kḱgǵ*
kḱgǵ + *x*	*pṕbb́* + *pṕbb́*	*td* + *cćčzźž*
cćčzźž + *x*	*fv́* + *pṕbb́*	

Невозможны в основном сочетания или идентичных, или различающихся на один признак консонантных классов, а также сочетание класса *cćčzźž* с *x*. Следует отметить, что не все одинаковые классы не могут встречаться рядом. Это связано с тем, что учитывались проклитические союзы и предлоги.

Встречаемость консонантных классов друг с другом

7 ранг	6 ранг	5 ранг	4 ранг	3 ранг	2 ранг	1 ранг	Консонантный класс
—	—	—	*pṕbḃ* 3	*td* 3	*sśšzźž* 4	*fḟvṽ* 6	*x*
—	*kḱgǵ* 10	*pṕbḃ* 22	*cćčzźž* 33	*fḟvṽ* 36	*td* 87	*sśšzźž* 113	*kḱgǵ*
pṕbḃ 4	*fḟvṽ* 5	*x* 9	*cćčzźž* 10	*kḱgǵ* 27	*td* 69	*sśšzźž* 106	*fḟvṽ*
—	—	*cćčzźž* 7	*kḱgǵ* 15	*fḟvṽ* 18	*td* 20	*sśšzźž* 126	*pṕbḃ*
cćčzźž 4	*x* 9	*sśšzźž* 23	*td* 40	*kḱgǵ* 79	*fḟvṽ* 103	*pṕbḃ* 183	*sśšzźž*
x 1	*cćčzźž* 6	*pṕbḃ* 8	*td* 9	*fḟvṽ* 40	*kḱgǵ* 113	*sśšzźž* 406	*td*
—	*x* 7	*pṕbḃ* 13	*kḱgǵ* 16	*cćčzźž* 20	*fḟvṽ* 25	*sśšzźž*	*cćčzźž*

В таблице сразу выделяется класс *x* тем, что у него наиболее ограниченная встречаемость — по три места не заполнены в α- и β-поле. С другой стороны, выделяются остальные классы непрерывных фонем *fḟ vṽ* и *sśšzźž* тем, что они обладают свойством логической симметричности относительно всех классов (и, следовательно, свойством рефлексивности) — они встречаются в любом сочетании со всеми консонантными классами. Класс *fḟ vṽ* помимо этого обладает и свойством ранговой симметричности в отношении классов первых четырех рангов в α- и β-поле. Уже по одному этому классы непрерывных фонем отделяются от других классов и друг от друга.

Соответственно этим же способом разделяются и классы прерывных фонем: классы *cćčzźž* и *kḱgǵ* объединяются тем, что у них заполнено одинаковое число мест в α-и β-поле. Оставшиеся классы *pṕbḃ* и *t d* различаются тем, что у первого больше мест занято в β-поле, а у второго — в α-поле.

Однако различить между собой классы *kḱgǵ* и *cćčzźž*, пользуясь лишь критериями симметричности и «полноты», уже невозможно. Надо рассмотреть, какие именно консонантные классы занимают то или иное место после данного класса.

Рассмотрим сначала «мощность» α-и β-полей консонантных классов (m_α и m_β — т. е. число встречаемостей в α-и β-поле). Здесь опять от всех отличается *x*, показывающий очень слабую встречаемость ($m_\alpha = 16$, $m_\beta = 26$). Отметим, что классы периферийных отличаются от классов непериферийных не только тем, что m_α и

Таблица 8

1 ранг	2 ранг	3 ранг	4 ранг	5 ранг	6 ранг	7 ранг
s ś š z ź ž 9	*f ƒ v v́* 9	*c ć č ź ǯ ǯ́* 7	*t d* 1	—	—	—
t d 113	*s ś š z ź ž* 79	*f ƒ v v́* 27	*c ć č ź ǯ ǯ́* 16	*p ṕ b b́* 15	*k ḱ g ǵ* 10	—
s ś š z ź ž 103	*t d* 40	*k ḱ g ǵ* 36	*c ć č ź ǯ ǯ́* 25	*p ṕ b b́* 18	*x* 6	*f ƒ v v́* 5
s ś š z ź ž 183	*k ḱ g ǵ* 23	*c ć č ź ǯ ǯ́* 13	*t d* 8	*f ƒ v v́* 4	*x* 3	—
t d 406	*c ć č ź ǯ ǯ́* 167	*p ṕ b b́* 126	*k ḱ g ǵ* 113	*f ƒ v v́* 106	*s ś š z ź ž* 23	*x* 4
k ḱ g ǵ 87	*f ƒ v v́* 69	*s ś š z ź ž* 40	*p ṕ b b́* 20	*t d* 9	*x* 3	—
k ḱ g ǵ 33	*c ć č ź ǯ ǯ́* 20	*f ƒ v v́* 10	*p ṕ b b́* 7	*t d* 6	*s ś š z ź ž* 4	—

m_β у непериферийных выше, чем у периферийных (подтверждение гипотезы Ципфа), но и тем, что у периферийных m_α не очень отличается от m_β, а для *f ƒ v v́* эти величины практически совпадают ($m_\alpha = 230$, $m_\beta = 233$). Во всяком случае, нигде не бывает превышения одной величины над другой вдвое. Иначе обстоит дело у непериферийных. Здесь одна величина превышает другую минимум в два раза (для *s ś š z ź ž* $m_\alpha = 441$, $m_\beta = 945$; для *t d* $m_\alpha = 582$, $m_\beta = 228$; для *c ć č ź ǯ ǯ́* $m_\alpha = 249$, $m_\beta = 80$).

Из общих закономерностей, характеризующих совместную встречаемость классов согласных, отметим следующие:

— положительное значение признака «периферийность» имплицирует положительное значение этого признака у консонантного класса, который чаще всего встречается за данным;

— отрицательное значение признака «периферийность» может быть связано как с положительным, так и с отрицательным значением этого признака у консонантного класса, который чаще всего следует за данным;

— равным образом в самых частых консонантных парах положительные и отрицательные значения признака «непрерывность—прерывность» могут сочетаться произвольным образом.

Вышесказанное относилось к сочетанию данного консонантного класса с последующим классом 1 ранга. Картина не изменяется, если рассмотреть все отношения прерывных—непрерывных и периферийных—непериферийных в окружении слева.

Для фонемы *x* — 17 непериферийных, 9 периферийных, 18 непрерывных, 8 прерывных

Для класса *kḱgǵ* — 208 непериферийных, 52 периферийных, 106 непрерывных, 154 прерывных

Для класса *ff́vv́* — 168 непериферийных, 65 периферийных, 114 непрерывных, 119 прерывных

Для класса *pṕbb́* — 204 непериферийных, 29 периферийных, 190 непрерывных, 43 прерывных

Для класса *sśšzźž* — 596 непериферийных, 349 периферийных, 812 прерывных, 133 непрерывных

Для класса *td* — 49 непериферийных, 179 периферийных, 112 непрерывных, 116 прерывных

Для класса *cćčzźž* — 30 непериферийных, 50 периферийных, 10 непрерывных, 66 прерывных

Как в случае сочетаемости с последующим классом 1-го ранга, так и при рассмотрении всего соотношения периферийных—непериферийных, единственный класс, не подчиняющийся правилу о том, что после класса, идентифицируемого данным значением признака «периферийность», чаще всего встречаются согласные с противоположным значением этого признака, является класс *sśšzźž*. Таким образом, по всем критериям статистической сочетаемости этот класс отделяется от всех других. Его особое положение и в однофонемной статистике (относительная стабильность частоты) — все это свидетельствует об особом статусе этой группы фонем в польском языке и подтверждает нашу мысль о том, что для польского языка этот класс типологически маркирован.

Сравнивая статистическое распределение встречаемости консонантных классов, мы можем выделить два класса — *t d* и *kḱgǵ*; распределение консонантных классов в соответствующих β-полях почти зеркальное. После *kḱgǵ* классом 1-го ранга является *t d*, после *t d* — *kḱgǵ*, т. е. прерывные фонемы с противоположным значением признака «периферийность»; затем после *kḱgǵ* следует *sśš zźž*, а после *t d* — *ff́ vv́*, т. е. непрерывные фонемы с противоположным значением «периферийности»; далее после *kḱgǵ* следует класс *ff́ vv́*, а после *t d* — *sśšzźž*, т. е. фонемы, относящиеся к тому же самому значению «периферийности», что и сами сравниваемые классы, но с противоположным значением признаков «непрерывность». И, наконец, на четвертом месте после *kḱgǵ* стоит *cćčzźž*, а после *t d* — *pṕbb́*, т. е. фонемы, идентифицируемые тем же значением «непрерывности», что и сравниваемые классы, но с противоположным значением «периферийности».

Класс *pṕbb́* отличается от остальных классов прерывных фонем тем, что после него на первом месте стоит класс *sśšzźž* При этом частотность *sśšzźž* после *pṕbb́* превышает частотность этого класса после любой другой группы согласных. Более того, эта величина (183) в общем соизмерима с частотностью *sśšzźž* после всех остальных групп согласных вместе взятых (258). Это явление можно со-

поставить с встречаемостью после *pṕbb* фонемы *r* (118 раз), что превышает встречаемость этой фонемы после любого другого класса согласных (после *kg* — 58, после *t d* — 91, после *f v* — 25, после *sśšzźž* — 21, после *cćčzźž* — 1).

Таким образом, класс *pṕbb* также получает специфический критерий для выделения его из других классов прерывных фонем.

Классы *sśšzźž* и *ff vv́* можно разделить, основываясь не только на ранговой симметричности класса *ff vv́* и на значительно большей мощности α-поля *sśšzźž* по сравнению с *ff vv́*, но и на том основании, что после *ff vv́* первое место занимает *sśšzźž* (непрерывный), а после *sśšzźž* — *t d* (прерывный).

Таковы основные черты статистической дистрибуции классов фонем в исследованных польских текстах. Общая тенденция распределения может быть сформулирована следующим образом: после консонантных классов статистически преобладают гласные, идентифицируемые тем же значением признака «периферийность—непериферийность», что и данный класс, и согласные, идентифицируемые противоположным значением этого признака (последнее не соблюдается для класса *sśšzźž*).

Как видим, даже самые предварительные результаты статистического исследования дистрибуции классов фонем оказываются весьма значимыми с фонологической точки зрения. Предложенная процедура анализа встречаемости с точки зрения правила Ципфа и в терминах α- и β-полей представляется приемлемой и открывающей новые перспективы в описании синтагматики. Дальнейшие исследования должны повторить описанное в настоящем разделе для других текстов с тем, чтобы проверить, насколько полученные выводы подтвердятся.

ЗАКЛЮЧЕНИЕ

Разумеется, в рамках одного исследования трудно было осветить все вопросы, связанные с проблематикой фонологической статистики. В частности, один из центральных вопросов, интересующих исследователей,— возможности типологии фонологических статистических структур — остался почти незатронутым. Это связано прежде всего с тем, что не существует работ по проверке однородности текстов относительно фонологических частот для других языков. Между тем возможности типологического исследования, связанные с изучением стабильности и нестабильности частот фонем, достаточно велики. Прежде всего, следует установить по самым различным языкам (в частности славянским, поскольку изучение начато с польского языка), какие фонемы и классы фонем тяготеют к стабильной или нестабильной частоте. В то время, как можно заранее предположить, что частоты гласных, несогласных и согласных во всех языках будут стабильны, стабильность частоты отдельных консонантных, вокалических и несогласных классов может оказаться самой различной. Могут выделиться так называемые специфические фонологические классы, чье присутствие в текстах будет постоянно определять общее звучание речи на данном языке. Таким образом, традиционная квантитативная типология, сравнивающая частоты идентичных классов фонем в разных языках, может быть углублена и усовершенствована.

Вторым направлением продолжения типологических исследований статистической структуры на фонологическом уровне может явиться сравнение функций распределения частот фонем в разных языках посредством критерия Смирнова. Здесь следует отметить, что дальнейшая разработка методики применения этого критерия к лингвистическим исследованиям требует углубления содержательной и статистической интерпретации этого критерия. Если в отношении критерия χ^2 при его использовании для распределения частоты фонем и классов фонем мы можем хотя бы приблизительно

указать, за счет чего увеличивается значение накопленного χ^2 и уменьшается значение q_i, то для критерия Смирнова этого пока сделать не удается.

Иными словами, мы еще недостаточно хорошо знаем, за счет чего реально может увеличиваться разница между значениями двух функций. Следовательно, надлежит исследовать ход изменений разности и посмотреть, на каких именно участках функции происходит накопление этих изменений. Тогда мы сможем определить, какие фонемы (или группы фонем) вызывают неоднородность по критерию Смирнова. По-видимому, на первых порах необходимо будет проводить перекрестные сравнения по нескольким выборкам из двух языков. Для критерия χ^2 мы знаем, что увеличение объема выборки в общем снижает шансы на однородность при сравнении частотных распределений фонем. Что же касается критерия Смирнова, то наши данные по польскому языку показывают бóльшую однородность для выборок одного языка. Из этого может следовать, что достаточно взять две произвольные (и даже не слишком большие) выборки из двух языков, чтобы по ним судить о близости функций распределения в двух языках в целом. Однако полное отсутствие экспериментальных данных не дает возможности полностью принять эту методику. Именно поэтому на настоящем этапе исследования следует сначала сравнить между собою несколько выборок в каждом языке в отдельности, а затем каждую из них — с выборками из другого языка. Если результаты подтвердят достаточность рассмотрения всего лишь пары выборок для составления исчерпывающего суждения о сравниваемых функциях распределения, то процедура установления близости этих функций будет значительно облегчена.

Третьим возможным подходом к типологии является сравнение заполнения частотной сетки различными фонемами в свете действия правила Ципфа. Этот подход дополняет изучение стабильных и нестабильных частот фонем изучением характера первичных и вторичных фонем. Кроме того, исследование частотных границ зон частых, средних и редких фонем в различных языкак также может оказываться плодотворным.

Но этим не исчерпываются эвристические возможности предлагаемой модели изучения статистической структуры. Эта модель может быть рассмотрена и для других уровней языка. Здесь основными будут вопросы выделения первичных (фонемы) и вторичных (классы фонем), а может быть и более сложных учетных единиц.

Завершая изложение нашего исследования, скажем, что построенная модель статистической структуры фонологического уровня — единая функция распределения частот фонем («лестница частот») для данного языка, разделяющаяся на участки, внутри которых порядок следования одних фонем определен правилом Ципфа, а для других является произвольным, из чего следует стабильность частот

одних и нестабильность частот других фонем, связанная помимо этого со стабильностью и нестабильностью частот различных дифференциальных признаков,— эта модель во многом опирается на пионерские работы Дж. К. Ципфа, а также рассматривается нами как подтверждение его идей. В заключение хочется еще раз указать на выдающуюся роль этого лингвиста, чьи замечательные труды снова начинают оказывать свое плодотворное воздействие на постановку и решение проблем фонологической статистики.

ПРИЛОЖЕНИЕ

Таблица абсолютных и относительных частот фонем в 1-й главе повести Я. Ивашкевича «Девушка и голуби»

Фонема	1-я зона	2-я зона	3-я зона	4-я зона	5-я зона
#	452—0,126	445—0,124	495—0,137	449—0,125	402—0,127
a	321—0,102	344—0,109	320—0,103	327—0,104	279—0,101
e	315—0,101	331—0,105	314—0,101	305—0,097	304—0,110
o	291—0,092	260—0,083	302—0,097	283—0,090	233—0,034
i	284—0,090	263—0,084	251—0,081	286—0,091	243—0,088
t	123—0,039	126—0,040	147—0,047	144—0,046	99—0,036
n	132—0,042	126—0,040	117—0,038	138—0,044	93—0,034
ɣ	92—0,029	104—0,033	95—0,031	104—0,033	116—0,042
u	97—0,031	89—0,028	121—0,039	101—0,032	83—0,030
r	113—0,036	97—0,031	90—0,029	101—0,032	77—0,028
v	103—0,033	93—0,029	55—0,018	69—0,022	52—0,019
s	104—0,033	88—0,028	93—0,030	87—0,028	81—0,030
p	87—0,028	86—0,027	85—0,027	93—0,029	96—0,035
k	98—0,031	87—0,028	99—0,032	77—0,024	82—0,029
m	78—0,025	85—0,027	94—0,030	77—0,024	84—0,030
d	84—0,027	77—0,025	77—0,025	79—0,025	75—0,027

(Окончание)

Фонема	1-я зона	2-я зона	3-я зона	4-я зона	5-я зона
j	68—0,021	84—0,027	101—0,032	73—0,023	59—0,021
ń	90—0,029	76—0,025	70—0,023	83—0,026	61—0,022
l	72—0,023	57—0,018	61—0,020	70—0,022	51—0,018
ś	52—0,016	56—0,018	64—0,021	56—0,018	65—0,024
š	45—0,014	63—0,020	58—0,019	55—0,017	59—0,021
z	64—0,020	64—0,020	43—0,014	44—0,014	50—0,018
g	35—0,011	54—0,017	51—0,016	57—0,018	47—0,017
ž	46—0,015	39—0,012	33—0,010	45—0,015	34—0,012
c	37—0,012	51—0,016	36—0,012	30—0,009	37—0,013
b	27—0,009	36—0,011	41—0,013	50—0,016	29—0,010
f	38—0,012	41—0,013	24—0,008	43—0,014	37—0,013
ć	33—0,010	38—0,012	40—0,013	30—0,09	28—0,010
č	34—0,011	32—0,010	32—0,010	37—0,012	30—0,011
x	49—0,015	23—0,007	29—0,009	37—0,012	28—0,010
v́	28—0,009	34—0,011	31—0,010	26—0,008	21—0,008
m	20—0,006	31—0,010	29—0,009	23—0,008	24—0,009
ʒ́	18—0,006	22—0,007	22—0,007	25—0,008	32—0,012
ḱ	15—0,005	28—0,009	21—0,007	31—0,010	19—0,007
ō	16—0,005	20—0,006	8—0,002	17—0,005	13—0,005
b́	8—0,002	19—0,006	17—0,006	14—0,004	11—0,004
ṕ	7—0,002	5—0,002	11—0,004	5—0,002	16—0,006
ʒ	7—0,002	6—0,002	6—0,002	11—0,004	6—0,002
ź	5—0,002	2—0,001	6—0,002	8—0,002	3—0,001
f́	3—0,001	3—0,001	4—0,001	3—0,001	4—0,001
ǵ	2—0,001	5—0,002	4—0,001	4—0,001	2—0,001
ē	4—0,001	3—0,001	3—0,001	3—0,001	3—0,002
ǯ	2—0,111	2—0	0—0	0—0	0—0

Таблица абсолютных и относительных частот фонем в 1-м акте пьесы Л. Кручковского «Первый день свободы»

Фонема	1-я зона	2-я зона	3-я зона	4-я зона	5-я зона	6-я зона	7-я зона
#	0,143—570	0,139—578	0,138—593	0,136—589	0,144—598	0,139—589	0,149—875
e	0,115—392	0,115—415	0,117—435	0,116—422	0,124—442	0,120—438	0,123—613
i	0,075—256	0,080—288	0,069—258	0,075—275	0,078—278	0,075—275	0,079—393
a	0,100—341	0,090—322	0,097—363	0,093—339	0,088—315	0,085—314	0,098—487
o	0,084—287	0,077—277	0,090—334	0,087—317	0,083—298	0,087—320	0,081—405
t	0,045—154	0,047—168	0,047—174	0,046—167	0,048—172	0,043—157	0,046—228
n	0,027—91	0,032—114	0,036—134	0,038—140	0,038—137	0,042—153	0,038—191
j	0,026—89	0,030—109	0,029—108	0,030—111	0,039—138	0,033—119	0,033—164
s	0,030—102	0,026—92	0,028—104	0,028—101	0,027—96	0,025—92	0,025—123
r	0,029—101	0,032—114	0,030—112	0,029—104	0,025—89	0,027—98	0,024—119
u	0,030—104	0,029—103	0,028—106	0,035—129	0,030—104	0,035—128	0,031—153
l	0,030—102	0,023—83	0,020—74	0,015—56	0,023—83	0,021—78	0,024—122
k	0,025—84	0,026—94	0,022—83	0,019—69	0,023—82	0,022—80	0,024—122
p	0,028—96	0,022—80	0,030—111	0,036—133	0,033—118	0,035—130	0,031—152
z	0,015—52	0,019—68	0,021—77	0,016—59	0,017—63	0,020—72	0,016—82
m	0,040—137	0,035—126	0,035—129	0,044—159	0,039—138	0,037—138	0,034—170
ń	0,025—84	0,031—110	0,028—105	0,032—116	0,026—94	0,029—106	0,032—160
v	0,020—68	0,027—95	0,023—84	0,021—77	0,022—79	0,023—86	0,020—99
d	0,023—79	0,020—72	0,023—86	0,016—57	0,017—58	0,016—60	0,021—107
ć	0,014—50	0,014—50	0,011—39	0,016—59	0,014—50	0,015—55	0,015—74
ɛ	0,012—43	0,014—50	0,011—40	0,018—66	0,011—41	0,012—42	0,012—61

(Окончание)

Фонема	1-я зона	2-я зона	3-я зона	4-я зона	5-я зона	6-я зона	7-я зона
g	0,013—50	0,017—60	0,011—42	0,010—38	0,012—42	0,012—43	0,012—61
ɣ	0,016—54	0,016—57	0,022—81	0,014—51	0,016—57	0,010—37	0,014—71
ś	0,021—73	0,021—76	0,018—69	0,015—55	0,022—78	0,021—76	0,027—132
ʒ́	0,006—18	0,008—28	0,012—48	0,011—42	0,012—42	0,012—42	0,011—52
ǰ	0,010—35	0,012—43	0,011—40	0,014—53	0,009—34	0,012—43	0,011—53
ć	0,020—70	0,021—76	0,023—84	0,018—65	0,018—65	0,021—77	0,018—88
š	0,024—83	0,025—91	0,022—81	0,023—82	0,023—84	0,033—122	0,019—95
ž	0,014—48	0,016—58	0,016—61	0,018—66	0,016—57	0,011—41	0,015—75
x	0,018—60	0,012—41	0,011—42	0,012—45	0,009—31	0,010—37	0,011—54
ʃ́	0,002—7	0,002—8	0,001—4	0,001—3	0,001—2	0,002—8	0,002—10
õ	0,005—19	0,008—31	0,006—22	0,006—23	0,008—29	0,007—27	0,005—26
ḱ	0,006—22	0,005—16	0,005—17	0,006—22	0,004—15	0,003—13	0,004—23
b	0,018—61	0,017—61	0,014—53	0,016—60	0,013—47	0,012—43	0,014—67
m̀	0,009—32	0,009—33	0,009—34	0,008—30	0,009—31	0,009—32	0,008—38
ʒ́	0,008—27	0,010—35	0,011—2	0,010—35	0,010—35	0,010—35	0,011—54
ǵ	0,001—2	0—0	0,001—2	0,001—2	0,001— 2	0,001—3	0,001—2
ʒ	0,003—9	0,001—5	0,003—9	0,001—5	0,004—15	0,003—10	0,003—15
b́	0,003—12	0,004—14	0,004—14	0,003—10	0,002—9	0,002—7	0,004—22
ź	0,002—4	0,001—4	0,002—6	0—1	0,001—2	0,002—7	0,001—5
ẽ	0—0	0,001—1	0,001—3	0—1	0,001—3	0,001—2	0,00—5
p̀	0,006—20	0,005—18	0,003—10	0,003—10	0,004—12	0,004—14	0,002—12
ǯ	0,001—1	0—0	0—0	0—1	0—0	0—1	0—1

Таблица абсолютных и относительных частот фонем в пяти рассказах С. Мрожека

Фонема	I Góral	II Interwał	III Nadzieja	IV Mały przyjaciel	V Jak walczyłem
#	730—0,130	989—0,147	1041—0,131	1086—0,126	1248—0,124
e	510—0,105	586—0,091	748—0,108	792—0,105	919—0,104
i	401—0,082	539—0,084	566—0,082	625—0,083	666—0,078
a	468—0,096	649—0,101	636—0,092	674—0,089	872—0,099
o	456—0,094	584—0,091	605—0,087	707—0,094	760—0,086
t	189—0,039	309—0,048	339—0,049	322—0,043	373—0,042
n	227—0,047	291—0,045	297—0,043	307—0,041	389—0,044
j	121—0,025	165—0,026	168—0,024	145—0,019	176—0,020
s	155—0,032	182—0,028	198—0,029	202—0,027	244—0,028
r	157—0,032	182—0,028	224—0,032	205—0,027	319—0,036
u	152—0,031	209—0,033	214—0,031	225—0,030	277—0,031
l	102—0,021	150—0,023	150—0,022	141—0,019	186—0,021
k	127—0,026	175—0,027	188—0,027	234—0,031	257—0,029
p	157—0,032	214—0,033	209—0,030	220—0,029	251—0,028
z	85—0,018	121—0,019	139—0,020	126—0,017	155—0,017
m	122—0,025	144—0,023	248—0,036	296—0,039	318—0,036
ń	127—0,026	175—0,027	185—0,027	208—0,027	194—0,022
v	141—0,029	197—0,031	169—0,024	210—0,028	235—0,027
d	119—0,024	148—0,023	171—0,024	185—0,024	235—0,027
c	79—0,016	70—0,011	77—0,011	73—0,010	126—0,014
č	45—0,009	82—0,013	96—0,014	104—0,014	94—0,011
g	78—0,016	89—0,014	80—0,012	136—0,018	122—0,014
ɣ	138—0,028	190—0,030	209—0,030	240—0,032	286—0,033
ś	78—0,016	131—0,021	126—0,018	151—0,020	160—0,018
ʒ́	48—0,010	57—0,009	60—0,009	70—0,009	72—0,008
f	62—0,013	59—0,009	84—0,012	86—0,012	107—0,012
ć	78—0,016	104—0,016	102—0,015	131—0,017	136—0,015
š	113—0,023	140—0,022	133—0,019	160—0,021	218—0,025
ž	63—0,013	68—0,010	97—0,014	135—0,018	108—0,012
x	61—0,012	56—0,009	58—0,008	94—0,012	88—0,010
ʃ́	6—0,001	12—0,002	15—0,002	12—0,001	22—0,003
ō	29—0,006	41—0,007	43—0,006	25—0,003	57—0,007
ǩ	34—0,007	39—0,006	48—0,007	42—0,006	38—0,004
b	58—0,012	82—0,018	98—0,014	122—0,016	126—0,014
m̀	26—0,005	44—0,007	50—0,007	48—0,006	75—0,009
ʒ́	23—0,004	46—0,007	23—0,003	42—0,005	62—0,007
ǵ	3—0,001	9—0,001	7—0,001	4—0	9—0,001
ʒ	5—0,001	12—0,002	18—0,003	10—0,001	13—0,002
b́	9—0,002	13—0,002	17—0,003	25—0,004	28—0,003
ź	6—0,001	13—0,002	7—0,001	12—0,001	7—0,001
ē	2—0,001	11—0,002	6—0,001	9—0,001	12—0,001
ṕ	12—0,002	27—0,004	23—0,003	12—0,001	23—0,003
ǯ	2—0,001	2—0	1—0	1—0	0—0

Таблица абсолютных и относительных частот фонем в пяти рассказах Ю. Шанявского

Фонема	I Profesor Tutka o złodzieju	II O słowie drukowanym	III Sprawa Osobista	IV O pszczołach i miodzie	V O dwóch malowidłach
#	592—0,134	815—0,135	909—0,138	956—0,136	917—0,125
e	462—0,121	571—0,110	917—0,109	627—0,101	715—0,111
i	300—0,079	411—0,079	498—0,088	530—0,085	520—0,081
a	376—0,099	517—0,099	496—0,088	569—0,091	589—0,092
o	312—0,082	471—0,091	504—0,089	540—0,087	567—0,088
t	153—0,040	205—0,040	286—0,050	263—0,042	300—0,047
n	160—0,042	193—0,037	225—0,040	228—0,037	252—0,039
j	119—0,031	127—0,025	161—0,030	129—0,021	160—0,025
s	101—0,027	149—0,029	177—0,031	181—0,029	198—0,031
r	124—0,033	146—0,028	207—0,036	163—0,026	231—0,036
u	114—0,030	190—0,037	163—0,028	224—0,036	221—0,034
l	93—0,024	115—0,022	130—0,023	135—0,022	100—0,015
k	109—0,029	170—0,032	151—0,027	165—0,026	164—0,026
p	143—0,038	179—0,035	196—0,035	200—0,032	214—0,033
z	56—0,014	72—0,014	94—0,017	120—0,020	107—0,017
m	145—0,038	172—0,033	222—0,039	208—0,033	241—0,037
ń	89—0,023	131—0,025	121—0,021	153—0,025	171—0,027
v	89—0,023	124—0,024	168—0,030	150—0,024	162—0,025
d	71—0,019	108—0,021	103—0,018	177—0,029	168—0,026
c	35—0,009	55—0,010	71—0,013	60—0,010	75—0,012
č	28—0,007	65—0,012	65—0,012	90—0,014	80—0,012
g	62—0,016	61—0,012	82—0,015	83—0,013	65—0,010
ɣ	101—0,027	162—0,031	133—0,023	187—0,030	195—0,030
ś	68—0,018	94—0,018	75—0,013	115—0,018	91—0,014
ɓ	33—0,009	74—0,014	66—0,011	66—0,010	71—0,011
ʃ	44—0,012	55—0,011	69—0,012	79—0,013	107—0,017
ɛ́	46—0,012	78—0,015	76—0,013	90—0,015	65—0,010
š	61—0,016	110—0,021	83—0,015	138—0,022	126—0,020
ž	80—0,021	77—0,015	92—0,016	100—0,016	93—0,014
x	40—0,010	31—0,006	64—0,011	67—0,011	64—0,010
ʃ́	8—0,002	12—0,002	12—0,002	13—0,002	13—0,002
ō	13—0,004	35—0,007	50—0,008	60—0,010	28—0,005
ǩ	24—0,006	25—0,005	25—0,004	39—0,007	30—0,005
b	46—0,012	51—0,010	85—0,015	71—0,012	65—0,010
ḿ	21—0,006	43—0,009	36—0,006	52—0,008	53—0,009
ʒ́	38—0,010	46—0,009	34—0,006	53—0,008	48—0,008
ɣ́	6—0,002	4—0,001	6—0,001	4—0,001	1—0
ʒ	9—0,002	11—0,002	7—0,001	9—0,001	8—0,001
ɓ́	12—0,003	16—0,003	17—0,003	15—0,002	17—0,003
ɛ́	7—0,002	5—0,001	5—0,001	12—0,002	12—0,002
ē	0—0	1—0	4—0,001	5—0,001	7—0,001
ṕ	8—0,002	33—0,006	7—0,001	23—0,004	20—0,003
ǯ	3—0,001	1—0	0—0	0—0	0—0

Таблица абсолютных и относительных частот классов фонем по 1-й главе повести Я. Ивашкевича «Девушка и голуби»

Классы фонем	1-я зона	2-я зона	3-я зона	4-я зона	5-я зона
Всего — 15319	3148	3149	3105	3151	2766
Гласные *a e i o u õ ẽ*	1328—0,422	1310—0,416	1319—0,424	1322—0,420	1158—0,419
Несогласные *m ḿ n ń r l ŭ j*	665—0,211	660—0,210	657—0,212	669—0,212	565—0,204
Носовые *m ḿ n ń*	320—0,102	318—0,101	310—0,100	321—0,102	262—0,095
Плавные *r l ŭ j*	345—0,109	342—0,109	347—0,112	348—0,110	303—0,109
Периф. согл. *k ḱ g ǵ p ṕ b b́ f f́ v v́*	500—0,159	514—0,163	472—0,152	509—0,161	444—0,160
Непериф. согл. *s ś š z ź ž t d c ć č ʒ ʒ́ ǯ*	654—0,208	665—0,211	657—0,212	651—0,207	599—0,217
Периф. комп. *k ḱ g ǵ x*	199—0,063	197—0,063	204—0,065	206—0,065	178—0,084
Периф. некомп. *p ṕ b b́ f f́ v v́*	301—0,096	317—0,100	268—0,087	303—0,096	266—0,096
Периф. комп. непрерывн. *x*	49—0,015	23—0,007	29—0,009	37—0,012	28—0,010
Периф. комп. прерывн. *k ḱ g ǵ*	150—0,048	174—0,056	175—0,056	169—0,053	150—0,056
Периф. некомп. непрерывн. *f f́ v v́*	172—0,055	171—0,054	114—0,037	141—0,045	114—0,041
Периф. некомп. прерывн. *p ṕ b b́*	129—0,041	146—0,046	154—0,050	162—0,051	152—0,055
Непериф. непрерывн. *s ś š z ź ž*	316—0,100	312—0,099	297—0,096	295—0,094	292—0,106
Непериф. прерывн. *t d c ć č ʒ ʒ́ ǯ*	338—0,108	353—0,112	360—0,116	356—0,113	307—0,111
Непериф. тускл. *t d*	207—0,1066	203—0,065	224—0,072	223—0,070	174—0,063
Непериф. яркие *c ć č ʒ ʒ́ ǯ*	131—0,042	150—0,047	136—0,044	133—0,043	133—0,048

Таблица абсолютных и относительных частот классов фонем по 1-му акту пьесы Л. Кручковского «Первый день свободы»

Классы фонем	1-я зона	2-я зона	3-я зона	4-я зона	5-я зона	6-я зона	7-я зона
Всего — 26589	3419	3584	3718	3656	3567	3659	4986
Гласные *a e i o u ō ē*	1399—0,409	1437—0,401	1518—0,408	1506—0,412	1469—0,412	1502—0,410	2082—0,418
Несогласные *m ḿ n ń r l ɫ j*	690—0,202	744—0,208	777—0,209	767—0,210	767—0,215	761—0,208	1035—0,207
Носовые *m ḿ n ń*	344—0,101	383—0,107	402—0,108	445—0,122	400—0,112	429—0,117	559—0,112
Плавные *r l ɫ j*	346—0,101	361—0,101	375—0,101	322—0,088	367—0,103	332—0,091	476—0,095
Периф. согл. *k ḱ g ǵ x p ṕ b b́ f f́ v v́*	535—0,157	558—0,156	550—0,148	564—0,153	515—0,144	549—0,150	729—0,146
Непериф. согл. *s ś š z ź ž t d c ć č ʒ ʒ́ ǯ*	795—0,232	845—0,235	873—0,235	819—0,225	816—0,229	847—0,232	1140—0,229
Периф. комп. *kḱgǵx*	218—0,064	211—0,050	186—0,050	176—0,048	172—0,048	176—0,048	262—0,052
Периф. некомп. *pṕbb́f f́*	317—0,093	347—0,096	364—0,098	388—0,105	343—0,096	373—0,102	467—0,094
Периф. некомп. непрерывн. *f f́ v v́*	128—0,038	174—0,048	176—0,047	175—0,047	157—0,044	179—0,049	214—0,043
Периф. некомп. прерывн. *pṕbb́*	189—0,055	173—0,048	188—0,051	213—0,058	186—0,052	194—0,053	179—0,049
Непериф. непрерывн. *s ś š z ź ž*	362—0,106	389—0,108	398—0,107	364—0,100	380—0,106	410—0,112	512—0,103
Непериф. прерывн. *t d c ć č ʒ ʒ́ ǯ*	433—0,126	456—0,127	475—0,128	455—0,125	436—0,123	437—0,120	628—0,126
Непериф. тускл. *t d*	233—0,068	240—0,067	260—0,070	224—0,062	230—0,064	217—0,059	335—0,067
Непериф. яркие *c ć č ʒ ʒ́ ǯ*	200—0,058	216—0,060	215—0,058	231—0,068	206—0,059	220—0,061	293—0,059

Таблица абсолютных и относительных частот классов фонем в рассказах Е. Шанявского

Классы фонем	Profesor Tutka o złodzieju	O słowie drukowanym	Sprawa osobista	O pczo łach i młodzie	O dwóch malowidłach
Всего — 27307	3809	5196	5683	6212	6414
Гласные *a e i o u õ ẽ*	1577—0,414	2196—0,422	2332—0,410	2555—0,411	2647—0,412
Несогласные *m n ḿ ń r l ų j*	852—0,224	1089—0,210	1235—0,217	1255—0,202	1403—0,219
Носовые *m n ḿ ń*	415—0,109	539—0,104	604—0,106	641—0,103	717—0,112
Плавные *r l ų j*	437—0,115	550—0,106	631—0,111	614—0,099	686—0,107
Периф. согл. *k ḱ g ǵ x p ṕ b b́ f f́ v v́*	624—0,164	835—0,160	948—0,167	975—0,157	993—0,155
Непериф. согл. *t d c ʒ ć ʒ́ č ǯ s ś š z ź ž*	756—0,198	1076—0,207	1168—0,206	1427—0,230	1371—0,214
Периф. комп. *k ḱ g ǵ x*	241—0,063	291—0,056	328—0,058	358—0,058	324—0,051
Периф. некомп. *p ṕ f f́ v v́ b b́*	383—0,101	544—0,105	620—0,109	617—0,099	669—0,104
Периф. комп. непрерывн. *x*	209—0,055	279—0,054	305—0,054	309—0,050	316—0,049
Периф. комп. прерывн. *k ḱ g ǵ*	174—0,046	265—0,051	315—0,055	308—0,049	353—0,055
Периф. некомп. прерывн. *p ṕ b b́*	400—0,010	31—0,006	64—0,011	67—0,011	64—0,011
Периф. некомп. прерывн. *f f́ v v́*	201—0,053	260—0,050	264—0,047	291—0,047	260—0,040
Непериф. непрерывн. *s ś š z ź ž*	373—0,098	507—0,098	526—0,093	666—0,107	627—0,098
Непериф. прерывн. *t d c ʒ ć ʒ́ č ǯ*	383—0,100	569—0,109	642—0,113	761—0,123	744—0,116
Непериф. прерывн. некомп. *t d c ʒ*	268—0,070	379—0,073	467—0,082	509—0,082	551—0,086
НепериФ. прерывн. комп. *ć ʒ́ č ǯ*	115—0,030	190—0,036	175—0,031	252—0,041	193—0,030
Непериф. неяркие *t d*	224—0,058	313—0,060	389—0,069	440—0,071	468—0,073
Непериф. яркие *c ć č ʒ ʒ́ ǯ*	159—0,042	256—0,049	253—0,044	321—0,052	276—0,043

Таблица абсолютных и относительных частот классов фонем в рассказах С. Мрожека

Классы фонем	I	II	III	IV	V
Всего — 34995	4874	6427	6931	7559	8315
Гласные *a e i o u õ ẽ*	2018—0,415	2629—0,409	2818—0,407	3057—0,405	3563—0,404
Несогласные *m n ḿ ń r l u̯ j*	1020—0,209	1341—0,209	1531—0,221	1590—0,210	1943—0,221
Носовые *m n ḿ ń*	502—0,103	654—0,102	780—0,113	859—0,113	976—0,111
Плавные *r l u̯ j*	518—0,106	687—0,107	751—0,108	731—0,097	967—0,110
Периф. согл. *k ḱ g ǵ x p ṕ b b́ f f́ v v̆*	796—0,163	1029—0,160	1056—0,152	1267—0,167	1378—0,156
Непериф. согл. *t d c ʒ ć ʒ́ č ǯ s ś š z ź ž*	1040—0,213	1428—0,222	1527—0,220	1645—0,218	1931—0,219
Периф. комп. *k ḱ g ǵ x*	303—0,062	368—0,057	381—0,055	510—0,067	514—0,058
Периф. некомп. *p ṕ b b́ v v́ f f́*	493—0,101	661—0,103	675—0,095	757—0,100	864—0,098
Периф. некомп. прерывн. *p ṕ b b́*	61—0,012	56—0,009	58—0,008	94—0,012	88—0,010
Периф. некомп. непрерывн. *f f́ v v́*	242—0,050	312—0,048	323—0,047	406—0,055	426—0,048
Периф. комп. непрерывн. *x*	236—0,048	336—0,052	347—0,050	379—0,050	428—0,048
Периф. комп. прерывн. *k ḱ g ǵ*	257—0,053	325—0,051	328—0,047	378—0,050	436—0,050
Непериф. непрерывн. *s ś š z ź ž*	500—0,103	655—0,102	700—0,101	786—0,104	892—0,101
Непериф. прерывн. *t d c ʒ ć ʒ́ č ǯ*	540—0,110	773—0,120	827—0,119	859—0,114	1039—0,118
Непериф. прерывн. некомп. *t d c ʒ*	392—0,080	539—0,084	605—0,087	590—0,078	747—0,085
Непериф. прерывн. комп. *c ʒ ǯ č*	148—0,030	234—0,036	222—0,032	278—0,036	292—0,033
Непериф. неяркие *t d*	308—0,063	457—0,071	510—0,073	507—0,067	608—0,069
Непериф. яркие *c ć č ʒ ʒ́ ǯ*	232—0,047	316—0,049	317—0,046	352—0,047	431—0,047

ЛИТЕРАТУРА

Общая фонология. Вопросы лингвистического моделирования

Ахманова О. С., Мельчук И. А., Падучева Е. В., Фрумкина Р. М. О точных методах исследования языка. М., 1961.

Бондарко Л. В., Зиндер Л. Р. Дифференциальные признаки фонем и их физические характеристики. «XIII Международный психологический конгресс. Москва, 1966. Симпозиум 23. Модели восприятия речи». Л., 1966.

Зиновьев А. А., Ревзин И. И. Логическая модель как средство научного исследования. «Вопросы философии», 1960, № 1.

Иванов В. Вс. О применимости фонологических моделей. «Труды ИТМ и ВТ АН СССР», вып. 2. М., 1961.

Лекомцева М. И., Сегал Д. М., Судник Т. Ш., Шур С. М. Опыт построения фонологической типологии близкородственных языков. «Славянское языкознание». М., 1963.

Лекомцева М. И. К описанию фонологической системы старославянского языка на основе тернарного принципа. «Лингвистические исследования по общей и славянской типологии». М., 1966.

Лиз Р. Б. О возможностях проверки лингвистических положений. «Вопросы языкознания», 1962, № 4.

Ревзин И. И. Модели языка. М., 1962.

Ревзин И. И. Некоторые вопросы теории моделей языка. «Научно-техническая информация», 1964, № 8.

Ревзин И. И. Метод моделирования и типология славянских языков. М., 1967.

Топоров В. Н. Предварительные материалы к описанию фонологических систем дардских языков. «Лингвистические исследования по общей и славянской типологии». М., 1966.

Трубецкой Н. С. Основы фонологии. М., 1960.

Шаумян С. К. Проблемы теоретической фонологии. М., 1962.

Шаумян С. К. Структурная лингвистика. М., 1965.

Якобсон Р. О. К характеристике евразийского языкового союза. Париж, 1931.

Chomsky N. On the notion «rule of grammar». «Proceedings of Symposia in Applied Mathematics», v. XI. Structure of language and its mathematical aspects. N. Y., 1961.

Halle M. The Sound Pattern of Russian. 's-Gravenhage, 1959.

Jakobson R., Halle M. Fundamentals of Language. 's-Gravenhage, 1956.

Kučera H. Inquiry into coexisting phonemic systems in Slavic languages. 's-Gravenhage, 1958.
Pilch H. Zentrale und periphere Lautsysteme. «Proceedings of the 5th International Congress of Phonetic Sciences». Basel — N. Y., 1965.
Vachek J. On peripheral phonemes. «Proceedings of the 5th International Congress of Phonetic Sciences». Basel — N. Y., 1965.

Лингвистическая статистика

1. Учебники и таблицы математической статистики

Большов Л. Н., Смирнов Н. В. Таблицы математической статистики. М., 1966.
Ван дер Варден Б. Л. Математическая статистика. М., 1960.
Янко Я. Математико-статистические таблицы. М., 1961.

2. Работы по лингвистической статистике общего характера

Гриднева Л. М. Розподіл голосних, приголосних і пропусків у сучасному українському мовленні. «Статистичні та структурні лінгвістичні моделі». Київ, 1966.
Лесскис Г. А. О зависимости между размером предложения и его структурой в разных видах текста. «Вопросы языкознания», 1964, № 3.
Лесскис Г. А. К вопросу о грамматических различиях научной и художественной прозы. «Труды по знаковым системам II». Тарту, 1966.
Лесскис Г. А. Два способа описания внеязыковых ситуаций. «Лингвистические исследования по общей и славянской типологии». М., 1966.
Марков А. А. Опыт статистического исследования романа «Евгений Онегин». «Известия Росс. имп. Академии наук». Серия 6, т. 7. СПб., 1913.
Савицкий Н. П. Об устойчивости относительных частот лингвистических элементов. «Ceskoslovenská rusistika», 1966, № 4.
Сегал Д. М. Некоторые уточнения вероятностной модели Ципфа. «Машинный перевод и прикладная лингвистика», № 5. М., 1961.
Сегал Д. М. Статистическая однородность текста на фонологическом уровне в польском языке. «Структурная типология языков». М., 1966.
Стойкова Л. С. До застосування вибіркового методу в лінгвістичних дослідженнях. «Статистичні та структурні лінгвистичні моделі». Київ, 1966.
Турыгина Л. А., Боркун М. Н. Статистические методы исследования частотного распределения лингвистических единиц. «Энтропия языка и статистика речи». Минск, 1966.
Фрумкина Р. М. О законах распределения слов и классов слов. «Структурно-типологические исследования». М., 1962.
Фрумкина Р. М. Статистические методы изучения лексики. М., 1964.

Abernathy R. Рец. на книгу: О. С. Ахманова, И. А. Мельчук, Е. В. Падучева и Р. М. Фрумкина. О точных методах исследования языка. «International Journal of American Linguistics», 1967, v. 33, № 1.
Ayer J. Chance. «Scientific American», 1965, № 11.
Ginneken van J. De statistiek in taalwetenschap. «De Nieuwe Taalgids», 1915, № 9, Groningen.
Ginneken van J. Benutzung der statistischen Methoden für die Sprachwissenschaft. «Indogermanisches Jahrbuch», 1920, v. 10.
Ginneken van J. Ras en Taal. «Verhandelingen der Koninglijken Akademie van wetenschapen te Amsterdam». Aft. Letterkunde, 1935, XXXVI.
Ginneken van J. De Ontwikkelingsgeschedenis van de systemen der menschelijke Taalklanken. Amsterdam, 1932.
Ginneken van J. De Oorzaken der taalveranderingen. Amsterdam, 1930.

Ginneken van J. La biologie et la base d'articulation. «Journal de psychologie», 1932, XXX.
Guiraud P. Bibliographie critique de la statistique linguistique. Utrecht. Anvers, 1954.
Guiraud P. Charactères statistiques du vocabulaire. Paris, 1954.
Guiraud P. Problèmes et méthodes de la statistique linguistique. Paris, 1960.
Herdan G. Language as Choice and Chance. Groningen, 1956.
Herdan G. Type-token Mathematics. The Hague, 1960.
Herdan G. The Calculus of Linguistic Observations. The Hague, 1962.
Herdan G. Quantitative Linguistics. London, 1964.
Herdan G. The Advanced Theory of Language as Choice and Chance. Berlin, 1966.
Mathesius V. La structure phonologique du lexique du tchèque moderne. «Travaux du cercle linguistique de Prague», I. Paris, 1929.
Mathesius V. Zum Problem der Belastungs- und Kombinationsfähigkeit der Phoneme. «Travaux du cercle linguistique de Prague», IV. Praha, 1931.
Newman E. B. Statistical methods in phonetics. «Manual of Phonetics», ed. L. Kaiser. Amsterdam, 1957.
Reed D. W. A statistical approach to quantitative linguistic analysis. «Word», 1949, № 5.
Ross A. S. Philological probability problems. «Journal of the Royal Statistical Society», ser. B.,12, 1950.
Trnka B. A phonological analysis of present-day standard English. «Práce z vědeckých ustavů», XXXVII. Praha, 1935.
Trnka B. K výstavbě fonologické statistiky. «Slovo a slovesnost», 1949, № 11.
Twaddell W. F. Combinations of consonants in stressed syllables in German. «Acta linguistica», Kopenhagen, 1939, № 1.
Vachek J. Poznámky k fonologii českého lexika. «Listy filologické», 1947, v. 67.
Williams C. B. A note on the statistical analysis of sentence-length as a criterion of literary style. «Biometrika», 1940, v. 31.
Yule G. U. The Statistical Study of Literary Vocabulary. Cambridge, 1944.
Zipf G. K. Relative Frequency as a Determinant of Phonetic Change. «Harvard Studies in Classical Philology». Cambridge, Mass., 1929, № 40.
Zipf G. K. Selected Studies of the Principle of Relative Frequency in Language. Cambridge, Mass., 1932.
Zipf G. K. The Psycho-biology of Language. Boston, 1935.
Zipf G. K. Human Behaviour and the Principle of Least Effort. N. Y., 1946.
Zipf G. K. Statistical methods and dynamic philology. «Language», 1937, v. 13.
Zipf G. K. Homogeneity and heterogeneity in language. «Psychological Records» 1938, № 2.
Zipf G. K. Phonometry, phonology and dynamic philology; an attempted synthesis. «American Speech», 1938, v. 13.
Zipf G. K. The psychology of language. «Encyclopedia of psychology». N. Y., 1946.

3. Подсчет частот фонологических и графемных элементов

Маринова М., Маринов Ас. Статистически изследвания на фонемите в българския книжовен език. «Български език», 1964, 2—3.
Перебейнос В. И. Частота и сочетаемость фонем современного украинского языка. Киев, 1964.
Bechardas Pandit Prabodh. Phonemic and Morphemic Frequencies of the Gujarati Language. Poona, 1965.
Bhagwat Shriram Vasudeo. Phonemic Frequencies in Marathi and their Relation to Devising a Speed-script. Poona, 1961.
Búrca de Séan. Irish phoneme frequencies. «Orbis», 1960, t. IX, № 2.

Denes P. B. On the statistics of spoken English. «Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung», Bd 17, H. 1. Berlin, 1964.
Dewey G. Relative Frequency of English speech sounds. Cambridge, 1925.
Dujardin F. Journal des connaissances usuelles. Paris, 1834.
Förstemann E. Numerische Lautverhältnisse im Griechischen, Lateinischen und Deutschen. «Zeitschrift für vergleichende Sprachforschung begr. von A. Kuhn», Bd 1. Göttingen, 1852.
Förstemann E. Lautbeziehungen des Griechischen, Lateinischen und Deutschen zum Sanskrit. «Zeitschrift für vergleichende Sprachforschung begr. von A. Kuhn», Bd 2, 1853.
Förstemann R. Numerische Lautverhältnisse in Griechischen Dialekten. «Zeitschrift für vergleichende Sprachforschung begr. von A. Kuhn», Bd 2, 1853.
Käding F. Häufigkeitswörterbuch der deutschen Sprache. Steiglitz b. Berlin, 1898.
Kerckhoffs A. La cryptographie militaire. Paris, 1883.
Sedlâcek Z. Základni studii k českému těsnopisu. I. Stanovení poměrů frekvenčních, iteračních a kombinačních v jazyce českem. «Těsnopisné Rozhledy». Praha, 1924.
Steffen M. Częstość występowania głosek polskich. «Biuletyn Polskiego Towarzystwa Językoznawczego», 1957, zesz. 16.
Thierry-Meg J. J. Structure phonologique du Française. Phonographie à pente unique. Nouveau systéme d'écriture abrégée. Paris, 1813.
Voelker C. H. A comparative study of investigations of phonetic dispersion in connected American speech. «Archives néerlandaises de phonetique expérimentale», 1937, v. 13.
Wang W. S.-Y., Crawford J. Frequency studies of English consonants. «Language and Speech», 1960, v. 3, part 3.
Weiss M. Über die relative Häufigkeit der Phoneme des Schwedischen. «Statistical Methods in Linguistics», 1961, № 1.
Whitney W. D. The proportional elements of English utterance. «Proceedings of the American Philological Association», 1874, v. 14.
Whitney W. D. On the comparative frequency of occurrence of the alphabetic elements in Sanskrit. «Oriental and Linguistic Studies», 2nd series. N. Y., 1874.
Whitney W. D. Sanskrit Grammar. Boston, 1896.

Работы по квантитативной фонологической типологии

Bourdon B. L'expression des émotions et des tendences dans le langage. Paris, 1892.
Greenberg J. H. The nature and uses of linguistic typologies. «International Journal of American Linguistics», v. XXIII, 1957, № 2.
Greenberg J. H. Essays in Linguistics. Chicago, 1957.
Kramský J. A quantitative typology of languages. «Language and Speech», 1959, v. 2.
Kramský J. Fonologické využití samohlaskových fonémat. «Linguistica Slovaca», IV—VI. Bratislava, 1946—1948.
Kramský J. On the quantitative phonemic analysis of English mono-and disyllables. «Časopis pro moderní filologie», 1956, v. 38.
Kramský J. A quantitative analysis of Italian mono- di- and trisyllabic words. «Travaux linguistiques de Prague». L'École de Prague d'aujourd'hui». Prague, 1964.
Kramský J. A phonological analysis of Persian monosyllables. «Archív orientalní», 1947, v. 16.
Kučera H. Entropy, redundancy and functional load in Russian and Czech. «American Contributions to the 5th International Congress of Slavists». The Hague, 1963.

Menzerath P., Meyer-Eppler W. Sprachtypologische Untersuchungen. I. «Studia Linguistica». Lund, 1950.
Menzerath P. Typology of languages. «Journal of the Acoustical Society of America», 1950, v. 22.
Menzerath P. Architektonik des deutschen Wortschatzes. Berlin, 1954.
Newman E. B. Pattern of vowels and consonants in various languages. «The American Journal of Psychology», 1951, v. 64.
Pierce J. E. A statistical study of consonants in New World languages. «International Journal of American Linguistics», 1957, v. XXIII.
Saporta Sol. Methodological considerations regarding a statistical approach to typologies. «International Journal of American Linguistics», 1957, v. XXIII.
Voegelin C. F. Inductively arrived-at models for cross-genetic comparison of American Indian languages. «University of California Publications in Linguistics», 1954, 10.
Wells R. Archiving and language typology. «International Journal of American Linguistics», 1954, v. 20.

Литература по польской фонологии

Толстая С. И. К типологической интерпретации польского ринезма. «Лингвистические исследования по общей и славянской типологии». М., 1966.
Шаумян С. К. История системы дифференциальных элементов в польском языке. М., 1959.
Benni T. Fonetika opisowa języka polskiego. Wrocław, 1959.
Biedrzycki L. Fonologiczna interpretacja polskich głosek nosowych. «Biuletyn Polskiego Towarzystwa Językoznawczego», z. 22, 1963.
Dłuska M. Fonetyka polska, c. I. Kraków, 1950.
Foliejewski Z. The problem of Polish phonems. «Scando-Slavica», t. II.
Jassem W. Рец. на книгу: M. Dłuska. Fonetyka polska, cz. I. Kraków, 1950, «Lingua posnaniensis», III. Poznań, 1951.
Jassem W. The distinctive features and the entropy of the Polish phoneme system. «Biuletyn Polskiego Towarzystwa Językoznawczego», z. 24, 1966.
Jassem W. A phonologic acoustic classification of Polish vowels. «Zeitschrift für Phonetik und allgemeine Sprachwissenschaft», Bd II, 1958, 4.
Materiały i prace Komisji Językowej, v. I. Warszawa, 1890—1891.
Milewski T. Derywacja fonologiczna. «Biuletyn Polskiego Towarzystwa Językoznawczego», z. 9, 1949.
Nitsch K. Stosunek [i] do [y]; spółgłoski podniebienne i niepodniebienne. «Wybór pism polonistycznych». Wrocław, 1954.
Nitsch K. Z historii rymów polskich. «Wybór pism polonistycznych». Wrocław, 1954.
Skorupka S. Studia nad budową akustyczną samogłosek polskich. Wrocław, 1955.
Stieber Z. Rozwój fonologiczny języka polskiego. Warszawa, 1962.
Stieber Z. O zaburzeniach równowagi fonologicznej. «Biuletyn Polskiego Towarzystwa Językoznawczego», z. 9, 1949.
Stankiewicz E. The phonemic pattern of Polish dialects. «For Roman Jakobson». The Hague, 1957.
Szober S. Gramatyka języka polskiego. Warszawa, 1959.
Zwoliński P. Dokoła fonemów potencjalnych. «Lingua Posnaniensis», IV. Poznań, 1951.
Zwoliński P. Stosunek fonemu [y] do [i] w historii językow słowiańskich. «Z polskich studiów sławistyczynch». Warszawa, 1958.

Литература по разным вопросам

Баринова Г. А. О произношении [к̄'] и [т̄']. «Развитие фонетики современного русского языка». М., 1966.

Падучева Е. В. О структуре абзаца. «Труды по знаковым системам II». Тарту, 1966.

Севбо И. П. Об изучении структуры связного текста. «Лингвистические исследования по общей и славянской типологии». М., 1966.

Сегал Д. М. О связи семантики текста с его формальной структурой «Poetics. Poetyka. Поэтика II». Warszawa, 1966.

Viet Jean. Les méthodes structuralistes dans les sciences sociales. Paris, 1965.

Zwirner E. Fonometrische Isophonen der Quantität der deutschen Mundarten. «Phonetica», 1959, № 4. Supplement.

ОГЛАВЛЕНИЕ

Дмитрий Михайлович Сегал

Основы фонологической статистики
(на материале польского языка)

Утверждено к печати
Институтом славяноведения
и балканистики АН СССР

Редактор издательства *Н. Н. Барская*
Художественный редактор *Т. П. Поленова*
Художник *Г. А. Астафьева*
Технический редактор *Е. Н. Евтянова*

Сдано в набор 15/X 1971 г.
Подписано к печати 7/I 1972 г. Формат 60×90¹/₁₆
Бумага № 2. Усл. печ. л. 16. Уч-изд. л. 16
Тираж 1400 экз. Тип. зак. 2950. Цена 96 коп.

Издательство «Наука»
Москва К-62, Подсосенский пер., 21
2-я типография издательства «Наука»
Москва Г-99, Шубинский пер., 10