

Мирослав Янакиев

ОСНОВЫ ТЕОРИИ ОРФОГРАФИИ*

Современный человек в основном усваивает орфографию в начальные годы своего пребывания в школе; орфография прочно закрепляется в его сознании и начинает казаться ему единственно возможной в своих «основных» (в смысле: «наиболее часто применяющихся») правилах. Из этой особенности орфографии вытекают два важных следствия.

Во-первых, каждая попытка усомниться в разумности основных орфографических правил расценивается всей национальной интеллигенцией как признак едва ли не потери здравого рассудка. Этим объясняется тот факт, что до сих пор так называемые «реформы» орфографии касались (и будут касаться впредь, по крайней мере еще некоторое время) фактически только мелочей, не затрагивая основных — врожденных или приобретенных — пороков ныне господствующих в мире орфографических систем. Во-вторых, большинство лингвистов фактически изучают не самые человеческие языки, а лишь их графические трансформации, которые приносят в речевое сообщение не свойственную ему «буквенную» дискретность и, таким образом, затушевывают его собственную дискретность.

Попытка проанализировать на строгой логической основе функционирующую в настоящее время систему орфографических понятий выявляет возможность вскрыть целый ряд неожиданных связей между усложнениями в научных описаниях звуковых языков и некоторыми особенностями орфографии. Поэтому построение строгой с логической точки зрения теории орфографии будет иметь значение не только для изучения существующих орфографий с целью их усовершенствования на научных основах, но и для решения некоторых до сих пор не решенных или неудовлетворительно решаемых общих проблем лингвистики и прежде всего фонологии. Построение логически строгой теории орфографии в первую очередь требует точного определения некоторых часто употребляемых терминов и разработки некоторой более точной системы терминов.

*

Будем рассматривать два непересекающихся множества: Φ — множество фактов, воспринимаемых слухом и называемых «речью»; Γ — множество фактов, воспринимающихся зрением и на-

* В сп. *Вопросы языкознания*, 1963, кн. 5, с. 47–57.

зываемых «буквенным письмом». И в фактах множества Φ , и в фактах множества Γ органы чувств человека четко разграничивают детали, которые в орфографиях не принимаются во внимание. Для теории орфографий элементарными фактами являются некоторые пересекающиеся множества элементарных фактов, представляющие собой подмножества множества Φ (соответственно множества Γ).

Число репрезентантов множества Γ в печатном тексте фактически определяется количеством появлений наиболее часто встречающегося элемента этого множества, который мы будем называть интерлитерой (межбуквенным пробелом) и будем обозначать в дальнейшем знаком Γ^1 . Например, число элементов множества Γ в цепочке *добрая* равно 6, потому что число интерлитер в ней равно 6: *д|о|б|р|а|я|*. Самое интерлитеру отдельным знаком обычно не считают, хотя с логической точки зрения таким образом допускается ошибка².

В цепочке, составленной из элементов множества Φ , нет элемента, характеризующегося фреквентностью, близкой к фреквентности интерлитеры. Этот факт вместе с непониманием роли интерлитеры как знака в письме весьма затрудняет попытки представить соответствия между элементами множества Φ и элементами множества Γ в форме, не вызывающей теоретических возражений.

Буквой называется сегмент письменного текста, как правило не содержащий интерлитер³.

Под «записыванием» и «чтением» надо понимать процессы трансформации элементов множества Φ в элементы множества Γ и наоборот, на основании каких-то правил однозначных соответствий, правил, совокупность которых и образует орфографию. Для установления однозначных соответствий нужно так сегменти-

-
- 1 Следует обратить внимание на кажущуюся очевидность числа вхождений (появлений) репрезентантов интерлитеры в данном печатном сообщении. Строго логический подход требует разграничить понятия «интерлитерема» (= класс эквивалентности интерлитер) и «интерлитера» (= конкретный, индивидуальный репрезентант, аллолитера «интерлитеремы»). Что касается рукописных текстов, то здесь следовало бы даже внести иерархию и классификации интерлитер — придется говорить о подклассе «подлинных» («чистых») интерлитер и о подклассе «штриховых» («лигатурных») интерлитер, так как в современных рукописных текстах буквы то отделяются друг от друга, то связываются наклонными штрихами («лигатурами»).
 - 2 Понятие «сегментации сообщения» предполагает понятие «сегментационного (демаркационного) сигнала». Правда, А. А. Сардинас и Дж. У. Паттерсон («Необходимое и достаточное условие однозначного разложения закодированных сообщений», «Кибернетический сборник», 3, М., 1961) предлагают другую процедуру однозначного разложения сообщений, якобы не требующую наличия специального сегментационного сигнала (пробела). Однако такая процедура невозможна, если предварительно не дана при помощи интерлитеры буквенная сегментация сообщения.
 - 3 Правда, в русской орфографии знак *ы* считается одной буквой, несмотря на наличие внутри него интерлитеры, однако это один из конкретно-исторически обусловленных алогизмов традиционной орфографии. Знак *ы* правильнее было бы рассматривать как дилитеру.

ровать речевое сообщение, чтобы для каждого его сегмента в его печатной письменной транскрипции имелся бы соответствующий графический сегмент (буква + интерлитера). Но так как в речевых сообщениях нельзя найти сигнала, появляющегося настолько часто, насколько часто появляется в письменных сообщениях интерлитера, то необходимо уменьшить каким-то образом частотность сегментирующих письменное сообщение интерлитер. Этого можно достичь, увеличивая разнообразие букв. Однако даже самый простой расчет показывает, что если сегментировать речевое сообщение на основании появления наиболее часто встречающегося в речи сигнала — вокальности, разнообразие букв для славянских языков должно было бы быть увеличено более чем в десять раз. Получился бы силлабарий, гораздо более сложный, чем японские.

Развитие письменности, как известно, всегда шло по пути ограничения разнообразия букв. Причины этого следует искать в более легком усвоении письма, характеризующегося меньшим разнообразием букв. Поэтому уменьшение частотности сегментирующих письменные сообщения интерлитер достигается иным путем. В современных орфографиях функционируют правила, на основании которых определенные цепочки букв трактуются как единичные буквы, т. е. сегментирующая функция интерлитеры в этих буквосочетаниях снимается. Ср., например, польск. *sz* или франц. *eau*. Вводом буквосочетаний, трактуемых как единичные буквы, и достигается то увеличение разнообразия письменных знаков, которое необходимо для реализации однозначного соответствия между сегментами письменного и сегментами речевого сообщения без увеличения разнообразия фигур букв, и, следовательно, без излишних затруднений процесса обучения письму.

Ввод в теорию орфографии понятия «буквосочетание, выполняющее функцию одной (единичной) буквы» требует ввода нескольких новых понятий. Во-первых, необходимо рассматривать межсловные пробелы как знаки, подобные обычным буквам. Весь раздел орфографии, регулирующий так называемые слитные и раздельные написания слов, является по существу списком орфографических правил употребления межсловных пробелов⁴. Назовем межсловный пробел термином «нуллолитера». Существует практика обозначать нуллолитеру (как аллолитеру нуллолитере-

4 Надо подчеркнуть, что трактовка межсловного пробела как орфографического знака обычна для прежних грамматических руководств (ср., например, *Л. В. Черепнин*, Русская палеография, М., 1956, стр. 375; *С. И. Абакумов*. Об основах методики пунктуации, «Изв. АПН РСФСР», 10, М.–Л., 1947, стр. 33). В этих руководствах межсловному пробелу даны и специальные названия. В данном отношении теория орфографии шла не вперед, а назад, и лишь развитие теории коммуникации (всеобщей теории связи) и ее раздела — теории информации — восстановило осознанность «знаковости» нуллолитеры.

мы) знаком]⁵. Во-вторых, теория орфографии требует отразить аналогичность структур буквосочетаний типа польск. *sz* и букв типа чеш. *š* (*š* орфографической точки зрения *š* не что иное, как сочетание *s* и *ˇ*, только *ˇ* ставится не справа от *s*, а над *s*). Следует выработать родовое понятие по отношению к понятиям «буква» и «диакритический знак», которые будут рассматриваться тогда как соподчиненные понятия. Репрезентанты этого понятия представляют собой минимальные объекты, для которых можно сформулировать орфографические правила. Будем называть каждый репрезентант этого понятия термином «литерон»⁶.

5 В. Penkov, А. Obretenov, В. Sendov, Т. Kirpikova, Т. Joukanov, Frequencies of letters in written Bulgarian, «Доклады Болгарской Академии наук», XV, 3, 1962, стр. 243–244.

6 Структура предлагаемого термина «литерон» не случайно аналогична структурам терминов типа «электрон», «протон», «фотон». Если рассматривать печатное письменное сообщение как цепочку дискретных позиций — литер, а каждую литеру (орфографическую позицию) как пространство (совокупность ячеек), каждой точке (ячейке) которого соответствует либо присутствие, либо отсутствие литерона определенного типа, то тогда печатное письменное сообщение будет интерпретацией теории цепей Маркова и к нему применима статистика Ферми–Дирака, играющая важную роль в современной статистической физике элементарных частиц.

Здесь невозможно подробно обсуждать далеко не маловажные вопросы о конструировании лингвистических и семиотических терминов, которым в лингвистике все еще не уделяют должного внимания. Необходимо подчеркнуть, однако, что применяемая здесь терминология является интерпретацией по отношению к орфографии формализованной терминологической системы, позволяющей преодолеть ряд хорошо известных затруднений, возникающих при описании языков. Изложим ее совсем сжато, только в виде соответствий с терминами теории множеств.

X — название признака, наличие которого у определенного объекта необходимо и достаточно, чтобы этот объект входил в многообразие, описываемое при помощи соответствующей терминологии;

X-он — наличие в данной позиции сообщения элемента m_i множества M ;

анти-*X*-он — отсутствие в данной позиции сообщения элемента m_i , множества M ;

X-ония — множество всех *X*-онов;

X-а — подмножество M_i множества M , характеризующее определенную позицию сообщения;

Y — название признака, наличие которого в данной *X*-е необходимо и достаточно, чтобы причислить *Y*-у к определенному классу, являющемуся подмножеством множества M ;

Y-*X*-а. — *X*-а класса *Y*;

Y-*X*-ия — класс, к которому принадлежит каждая *X*-а, характеризующаяся признаком *Y*;

Y-*X*-он — *X*-он, входящий в состав по крайней мере одной *Y*-*X*-ы;

Y-*X*-ония — множество всех *Y*-*X*-онов;

X-ия — множество \mathfrak{M} всех подмножеств M_i множества M ;

X-емия — подмножество \mathfrak{M}_i , множества \mathfrak{M} , объединение некоторых подмножеств M_1, M_2, \dots, M_n множества M с инвариантом, т. е. с непустым пересечением;

X-ема — непустое пересечение некоторых множеств M_1, M_2, \dots, M_n множества M ($inv \mathfrak{M}_i = \bigcap_{i=1}^n M_i \neq \emptyset$) или, что то же самое, инвариант какого-либо подмножества \mathfrak{M}_i множества \mathfrak{M} ;

Z — название *X*-емы;

алло-*X*-а *X*-емы — *Z*-репрезентант *X*-емы *Z* = реализация *X*-емы *Z* (можно обозначить и *Z*-*X*-а, если отвлечься от опасности недоразумения, порождаемого смешением с *Y*-*X*-ой) — подмножество M_i множества M , содержащее инвариант, называемый *X*-емой *Z*;

Z-*X*-емон — элемент m_i множества M , являющийся также элементом *X*-емы *Z*;

архи-*X*-ема *X*-ем P, Q, R, \dots — непустое пересечение некоторых *X*-ем P, Q, R, \dots ;

архи-*X*-емия *X*-емий P, Q, R, \dots — объединение *X*-ем P, Q, R, \dots

В-третьих, необходим термин «литера», которым будет обозначаться либо «буква», либо «буквосочетание», либо «буква с диакритическим знаком», либо, наконец, «сочетание, состоящее из букв и диакритических знаков», т. е. любая совокупность литеронов, для которой может быть сформулировано орфографическое правило. Термином «литера» будем называть именно такой сегмент печатного письменного текста, который читается и/или записывается на основании какого-то орфографического правила⁷.

При классификации литеронов и литер удобной является следующая терминология:

- «нуклеолитерон» — обычная буква, в том числе знаки препинания и дефис (но не диакритический знак);
- «перилитерон» — диакритический знак⁸;
- «супралитерон» — надстрочный диакритический знак, например, ˇ в ě или ˘ в ā;
- «сублитерон» — подстрочный диакритический знак, например, ˙ в ě (франц. *cédille*) или ˛ в польск. *ę*;
- «инлитерон» — строчный диакритический знак, например, ˚ в польск. *ł* или / в дат. *ø*;
- «нуклеолитера» — литера, состоящая только из нуклеолитеронов, например, *a*, *ы*, польск. *cz*, франц. *eau*, англ. *a—e*;
- «макролитерон» — перилитерон-конструкт, позволяющий рассматривать прописные буквы как дилитеры, образуемые из соответствующих строчных букв (нуклеолитеронов) и макролитерона; обозначим его знаком ⁹;
- «макролитера» — прописная буква, т. е. дилитера, состоящая из нуклеолитерона и макролитерона;
- «нуллолитерон» — литерон, образующий нуллолитеру; нуллолитера является также нуклеолитерой;
- «унилитера» — литера, состоящая из одного нуклеолитерона, но не нуллолитера;
- «монолитера» — нуллолитера либо унилитера;

7 Именно литере (а не «букве» в обычном смысле этого слова) соответствует в хорошо излагаемой Н. И. Жинкиным теории (см. «Знаки и система языка», сб. «*Zeichen und System der Sprache*», I, Berlin, 1961, а также «Звуковая коммуникация обезьян», «Изв. АПН РСФСР», 113 — Мышление и речь, М., 1960) термин «конкретная буква», или, точнее, «репрезентант какой-то абстрактной буквы». Термину «абстрактная буква», следуя привившейся уже в лингвистике удачной «аллоэмической» структуре терминов, соответствовал бы термин «литерема», а термину «репрезентант определенной абстрактной буквы» соответствовал бы термин «аллолитера определенной литеремы».

8 Некоторые перилитероны не отделены от нуклеолитерона явной интерлитерой, однако их функция требует, чтобы они рассматривались как отдельные орфографические объекты. Впрочем, явная интерлитера исчезает в определенных случаях и между двумя нуклеолитеронами. Ср., например, дат. и норв. *æ*. Вопрос о трактовке редуцированной до нуля интерлитеры с теоретической точки зрения интересен, но здесь нет необходимости его ставить.

9 Таким образом рассматривает прописные буквы, например, Г. Глисон («Введение в дескриптивную лингвистику», М., 1959, стр. 399).

- «полилитера» — сочетание литеронов, из которых по крайней мере один является нуклеолитероном, точнее унитероном (смежность нуллолитеронов в полилитере не допускается);
- «дилитера» («2-литера») — полилитера, состоящая из двух литеронов;
- «трилитера» («3-литера») — полилитера, состоящая из трех литеронов, причем если все они нуклеолитероны, первый и третий могут быть нуллолитеронами;
- «тетралитера» («4-литера») — полилитера, состоящая из четырех литеронов; аналогичным образом определяются термины «пенталитера» («5-литера»), «гексалитера» («6-литера») и т. д.

Полилитера может быть также «контактной» или «слитной», если между ее литеронами интерлитеры редуцированы до нуля (например, дат. *a*, *ø*, польск. *ł*, чеш. *č*, *á*), «недиссоциированной», если она контактная или если между ее литеронами «вставлены» только интерлитеры (например, *â*, *ë*, польск. *cz*, *ść*, *]]*) и «диссоциированной», если она состоит из литеронов, между которыми вставлены литероны других литер. Места диссоциаций в диссоциированной полилитере будем обозначать знаком —. Например, в английском слове *nate* сочетание *a—e* является диссоциированной дилитерой, так как оно состоит из двух литеронов — *a* и *e*, которые разделены (диссоциированы) литероном, принадлежащим унитеру *t*.

Следует иметь в виду, что в письменном сообщении одна и та же буква может быть одновременно репрезентантом литеронов, принадлежащих двум литеремам. Такую букву будем называть «диплолитероном». Например, в английском слове *age* буква *e* является в одно и то же время литероном диссоциированной дилитеремы *a—e* и литероном недиссоциированной дилитеремы *ge*.

Наконец, литеремы, объединяемые каким-то орфографическим правилом одинаковым способом чтения, будем называть «изоанагностическими» (от греч. основы ἰσoγvώστ-, соответствующей основе глагола «читать»), Изоанагностическими являются, например, в русской, украинской, белорусской и болгарской орфографиях дилитеремы *б?* и *п?*, *б!* и *п!*, *б*; и *п*;, *б*: и *п*: (но не *б*. и *п.*, поскольку они «псевдолитеремы» — об этом термине см. ниже — и их чтение может быть разным; ср., например, чтение *б*. в библиографическом сокращении *б. г.* [= «без года»] и в заканчивающем какое-то предложение слове *хлеб*; псевдолитеремами являются также *б][* и *п][*).

Литеры даже в печатном сообщении не вполне одинаковы. Две литеры являются одинаковыми, если их литеронный состав один и тот же, хотя реализации составляющих их литеронов могут различаться в деталях заметно для глаза, но случайно. Так, например, в выражении *Суд!* *Суд!* первая дилитера *д!* и вторая дилитера *д!* не вполне одинаковы — вершина литерона *!* во второй дилитере *д!* как бы отсечена. Однако всеми это отклонение в форме *д!*, несомненно, призналось бы «случайным». Литеры, одинаковые с точностью до литеронного изоморфизма, будем называть аллолитерами одной и той же литеремы.

Если принять, что в речевых сообщениях данного языка встречается во всяческих комбинациях f разных речевых фактов, не являющихся для общества фонетической особенностью данного говорящего индивида, то минимальное число литерем орфографии этого языка должно быть равным f , а число (l) литеронов, из которых должны конструироваться все f литерем орфографии этого языка, будет на единицу больше характеристики двоичного логарифма числа f . Для болгарского языка f не меньше 1194, а для русского языка f не меньше 1051¹⁰. Имеются в виду не только учитываемые в орфографических пособиях «фонемные» различия, но и такие «суперсегментные» дифференциации, как словесное ударение, фразовое ударение и интонация. Их подробное рассмотрение не может быть задачей настоящей статьи¹¹. Здесь дадим лишь общий их перечень для русского и болгарского звуковых языков.

Во-первых, это речевые факты, обозначаемые обычно сочетаниями согласной буквы со следующей за ней гласной. Всего их в русском звуковом языке 988. Они образуются из 38 консонантных фонем (19 твердых — *б, в, г, д, ж, з, к, л, м, н, п, р, с, т, ф, х, ц, ш* — и 19 мягких — *б', в', г', д', з', й, к', л', м', н', п', р', с', т', ф', х', ч, ц, ш*), и 26 вокальных фонем — 5 ударных (*а, о, у, э, ы*), 5 фразово (логически) акцентированных (*а, о, у, э, ы*) и 3 безударных (*а, у, ы*), каждая из которых может произноситься с восходящей либо с нисходящей интонацией. В болгарском языке соответствующих речевых фактов 1140. Они образуются из 39 консонантных фонем (22 твердых — *б, в, г, д, ж, з, й, к, л, м, н, п, р, с, т, ф, х, ц, ш, ч, ц, ш* — и 17 мягких — *б', в', г', д', з', к', л', м', н', п', р', с', т', ф'*,

10 Не меньше, но и не на много больше. Не включены в суммы лишь болгарская «инспираторная» (точнее «всасывающе» произносимая) альвеолярная аффриката, выражающая отрицание (при повторении — удивление), а также всасывающий билабиал «воздушного поцелуя», назальные ларингалы, выражающие непонимание, а при повторении согласие или несогласие, и, наконец, билабиальный вибрант, которым останавливают лошадей или подгоняют овец.

11 Словесное ударение, фразовое ударение, а также восходящая и нисходящая интонации ничем не отличаются от остальных «несуперсегментных дифференциальных признаков» и, по-видимому, не являются результатом, как обычно думают, деятельности голосовых связок, потому что различаются и при шепоте.

x' , u' , s'^{12}) и 30 вокальных — 6 ударных (a , e , u , o , y , \bar{z}), 6 фразово акцентированных (a , e , u , o , y , \bar{z}) и 3 безударных (a/\bar{z} , o/y , e/u), каждая из которых, как и в русском, может сопровождаться либо восходящей, либо нисходящей интонацией.

Во-вторых, в русском языке встречаются вне перечисленных сочетаний 26 вокальных (безударных — 3, ударных и фразово акцентированных по 5 с восходящей и нисходящей интонацией) и 36 консонантных фонем (включая звонкое x , но исключая мягкие z' , k' , x'). В болгарском языке — 30 вокальных фонем (безударных — 3, ударных и фразово акцентированных по 6 с восходящей и нисходящей интонацией) и 23 консонантные фонемы (включая звонкое x , но исключая все мягкие консонанты). Наконец, надо учесть и паузу.

Следовательно, число литеронов, необходимых и достаточных для выражения всего многообразия болгарского и русского звуковых языков (l), равно 11 ($10 < \lg_2 1051 < \lg_2 1194 < 11$) — имеются в виду литероны как детали букв с фиксированным местом.

А букв, т. е. литеронов, используемых в русском алфавите — 47, включая \ddot{e} , нуллолитеру, макролитерон, 12 знаков препинания. В болгарском алфавите их 44 — отсутствуют \ddot{e} , ε и \bar{y} .

Можно ли на основании этих данных утверждать, что русская и болгарская орфография являются неэкономными? Нелегко ответить коротко на этот вопрос. В максимально экономной 11-литеронной орфографии должны применяться нуллолитерема, 11 унилитерем, $55 (= C_{11}^2)$ дилитерем, $165 (= C_{11}^3)$ трилитерем, $330 (= C_{11}^4)$ тетралитерем, $462 (= C_{11}^5)$ пенталитеремы и из всех также 462 возможных ($= C_{11}^6$) гексалитерем только 27 для русского и 170 для болгарского языка. В буквах, применяющихся сейчас, распределение вариативных компонентов едва ли позволит так графически оформить литероны, чтобы новые литеремы выглядели похожими на старые буквы с их фонетическими значениями, к которым мы «безнадежно» привыкли. Лишь если бы это удалось, преимущества 11-литеронного алфавита стали бы очевидными. Вот почему нельзя ставить вопрос об оптимальной орфографии в собственно теоретическом плане.

Итак, множество $\Gamma = \{\Gamma_1, \Gamma_2, \dots, \Gamma_n\}$ является множеством не букв, а литерем, т. е. множеством множеств литеронов. Если представить себе такое разбиение множества Φ , что каждой литереме Γ_i множества Γ соответствует один и только один класс Φ_i разбиения множества Φ , то декартово произведение множеств Γ и Φ ($\mathfrak{D}_\Gamma = \Gamma \times \Phi$) можно назвать «орфоанагностией», а каждую пару — элемент декартова произведения множеств Γ и Φ ($\mathfrak{D}_{\Gamma ij} = \Gamma_i \Phi_j$) —

12 Перед e и u в болгарском звуковом языке твердые z , k , x не встречаются.

«элементарным правилом чтения» или «орфоанагностемой». Орфоанагностия не имеет отношения к возможности существования двух или более литерем ($\Gamma_i, \Gamma_k, \dots$) в качестве соответствий одного и того же класса Φ_j разбиения множества Φ , Короче:

$$\mathfrak{D}_\Gamma = \{ \mathfrak{D}_{\Gamma_{ij}} / \mathfrak{D}_{\Gamma_{ij}} = (\Gamma_i \Phi_j) \in \mathfrak{D}_\Gamma \Rightarrow \mathfrak{D}_{kj} = (\Gamma_k \Phi_j) \notin \mathfrak{D}_\Gamma, \text{ если } i \neq k \}. \quad (1)$$

В более привычной лингвистической терминологии определению (1) соответствует утверждение, что орфоанагностией обеспечивается отсутствие омографов в достаточно длинном тексте. Орфоанагностия делает возможным безошибочное одинаковое чтение (изоанагностия) текста¹³, но не имеет отношения к записыванию речи. Орфоанагностия призвана обеспечить чтение всех возможных текстов языка людьми, не имеющими никакого представления о семантике читаемых ими текстов.

Наоборот, если представить такое разбиение множества Γ , что некоторым подмножествам Φ_j, Φ_k, \dots множества Φ соответствует одна и только одна литераема Γ_i , множества Γ , то декартово произведение множеств Γ и Φ ($\mathfrak{D}_\Phi = \Phi \times \Gamma$) можно назвать «орфоанаграфией», а каждую пару декартова произведения множеств Φ и Γ ($\mathfrak{D}_{\Phi_{ij}} = (\Gamma_i, \Phi_j)$) — «элементарным правилом записывания» или «орфоанаграфемой». Орфоанаграфия не имеет отношения к возможности существования двух или более разных речевых фактов (Φ_j, Φ_k, \dots) в качестве соответствий одной и той же литеремы Γ_i , множества Γ . Короче, если обозначим орфоанаграфию через \mathfrak{D}_Φ , то:

$$\mathfrak{D}_\Phi = \{ \mathfrak{D}_{\Phi_{ij}} / \mathfrak{D}_{\Phi_{ij}} = (\Gamma_i \Phi_j) \in \mathfrak{D}_\Phi \Rightarrow \mathfrak{D}_{ik} = (\Gamma_i \Phi_k) \notin \mathfrak{D}_\Phi, \text{ если } i \neq k \}. \quad (2)$$

Иными словами, орфоанаграфия обеспечивает отсутствие омофонов в речевых сообщениях. Она делает возможным безошибочное одинаковое печатное отображение речевого сообщения, но не обеспечивает одинакового чтения записанного текста.

Всякая орфография ($\mathfrak{D}_{\Gamma\Phi}$) является пересечением орфоанагностии и орфоанаграфии ($\mathfrak{D}_{\Gamma\Phi} = \mathfrak{D}_\Gamma \cap \mathfrak{D}_\Phi$), т. е.

$$\mathfrak{D}_{\Gamma\Phi} = \left\{ \begin{array}{l} \mathfrak{D}_{\Gamma\Phi_{ij}} / \mathfrak{D}_{\Gamma\Phi_{ij}} = (\Gamma_i \Phi_j) \in \mathfrak{D}_{\Gamma\Phi} \\ \mathfrak{D}_{\Gamma\Phi_{ij}} / \mathfrak{D}_{\Gamma\Phi_{ij}} = (\Gamma_i \Phi_j) \in \mathfrak{D}_{\Gamma\Phi} \end{array} \right. \quad (3)$$

Ведь орфография, объединяющая правила записывания и правила чтения в правила об однозначных трансформациях (орфогра-

¹³ Под «безошибочным чтением» текста понимается не такое чтение, которое полностью воспроизводило бы первоначальную звуковую форму записанного сообщения, а такая звуковая интерпретация текста, все отклонения которой от первоначальной его звуковой формы воспринимаются обществом как допустимые индивидуальные особенности чтеца.

фемы), не должна содержать логически противоречивых утверждений.

Поэтому пары $\mathfrak{D}_{ij} = \{G_i\Phi_j\}$, $\mathfrak{D}_{kj} = \{G_k\Phi_j\}$ и $\mathfrak{D}_{ik} = \{G_i\Phi_k\}$ не могут принадлежать орфографии $\mathfrak{D}_{\Gamma\Phi}$. Иными словами, если наряду с правилом $\mathfrak{D}_{ij} = \{G_i\Phi_j\}$ существует правило $\mathfrak{D}_{kj} = \{G_k\Phi_j\}$, то при помощи этих правил можно читать, но для записывания речи они не годятся, поскольку противоречат друг другу. Обратно, если наряду с правилом $\mathfrak{D}_{ij} = \{G_i\Phi_j\}$ существует и правило $\mathfrak{D}_{ik} = \{G_i\Phi_k\}$, при их помощи можно записывать речь, но читать безошибочно невозможно. Например, на основании орфографических правил современной русской и болгарской орфографий «дилитереме *n?* соответствует неозвученный голосовыми связками экспираторный щелчок, производимый губами, за которым следует пауза» и «дилитереме *б?* соответствует такой же щелчок, за которым следует пауза» читать можно, но записывать нельзя, поскольку одна и та же последовательность речевых явлений должна была бы на основании первого соответствия трансформироваться в *n?*, а на основании второго соответствия — в *б?*. Наоборот, на основании, например, правила русской орфографии «дилитереме *no* соответствует сочетание, состоящее из производимого губами щелчка, за которым следует округленный широкий средний ударный вокал» и правила той же русской орфографии «дилитереме *no* соответствует сочетание, состоящее из производимого губами щелчка, за которым следует неокругленный средний безударный вокал» можно записывать, но читать нельзя.

Противоречия, порождаемые на основе объединения орфоаналогизмов и орфоанаграфем в одну систему утверждений, в орфографии преодолеваются путем применения орфоаналогизмов и орфоанаграфем, относящихся к полилитеремам, содержащим больше литеронов, чем отбрасываемые противоречащие друг другу правила. Так, например, если приведенные выше правила о дилитеремах *n?* и *б?* нельзя объединить в одну орфографию, то придется анализировать пары трилитерем, в которые входят дилитеремы *n?* и *б?*. Если при этом окажется, что среди этих трилитеремных пар есть такие, речевые компоненты которых не совпадают, и, кроме того, с их речевыми компонентами в данной орфографии не связаны в пары другие литеремы, то для соответствующих трилитерем формулируются орфографемы (взаимно однозначные орфографические правила трансформаций).

Таким образом, анаграфическая неопределенность дилитерем снимается, но зато вместо каждого дилитеремного правила в данной орфографии появляется столько трилитеремных правил, сколько литеронов применяется в орфографии. А если среди три-

литерем окажутся анаграфически или анагностически неопределенные, то необходимо будет перейти к тетралитеремам. И это снова теоретически умножит количество правил орфографии для каждой неопределенной трилитеремы во столько раз, сколько литеронов применяется в орфографии. То же произойдет с каждым переходом к пенталитеремам, гексалитеремам и т. д.

Так, например, анаграфическую неопределенность дилитерем $n?$ и $b?$ в русской и болгарской орфографиях невозможно снять ни трилитеремными, ни тетралитеремными правилами, и лишь некоторые пенталитеремы, оканчивающиеся дилитеремами $n?$ или $b?$, и анаграфически, и анагностически точно определены: пенталитеремы $][rab!$, $][zub!$, $][tin!$ и др. А количество всех орфографем в русской орфографии, относящихся к пенталитеремам, которые оканчиваются на $b!$ или $n!$, огромное: $2 \cdot 34^3 = 78\ 608$. И в болгарской орфографии оно того же порядка: $2 \cdot 31^3 = 59\ 582$. При этом не принимаются во внимание знаки препинания. Кроме того, даже некоторые пенталитеремы как для русской, так и для болгарской орфографии оказываются анаграфически неопределенными. Такими являются, например, русские $][tub!$ и $][tyn!$ и болгарские $][kub!$ и $][kyn!$.

Если вообще возможно составление перечней орфографических правил, то лишь потому, что в перечни включаются только полилитеремы, вероятность появления которых в текстах данного языка не слишком мала, т. е. только полилитеремы, которые в текстах данного языка отмечены средней фреквентностью, превышающей некоторую малую пороговую величину h . Орфографические словари представляют собой именно такие перечни. Право решения включить или не включить какую-либо полилитерему в словарь, т. е. считать эту полилитерему отмеченной принадлежностью или непринадлежностью к данному языку, предоставляется не очень надежным накапливающим памятным устройствам мозга лингвиста, составителя словаря. Лингвисты же из всех полилитерем, отмеченных вероятностью, большей h (названной ими «осмысленность»), признают достойными орфографического уважения лишь те, которые начинаются нуллолитероном и заканчиваются знаком препинания, обозначающим паузу; они заменяют этот знак нуллолитероном, называют соответствующие полилитеремы «словами» и объявляют их «кандидатами в словарь». Однако прежде чем включить слова (т. е. полилитеремы типа $][ab \dots n][$) в орфографический словарь, их подвергают еще одной «чистке» — среди «кандидатов» выбирают так называемые «основные формы» и лишь им дают «пропуск» в словарь. Орфографию остальных, «неоснов-

ных», хотя и более часто встречаемых в текстах форм слов орфографический словарь не дает.

Замена конечного знака препинания нуллолитероном представляет собой объяснимую (нуллолитереме может соответствовать в речи пауза), но с логической точки зрения опасную трансформацию, которая не замедляет дать о себе знать: некоторые включаемые в словари полилитеремы типа $][ab \dots n][$ являются анагностически и/или анаграфически неопределенными, поскольку их трансформы ($][ab \dots n:$, $][ab \dots n?$, $][ab \dots n!$ и т. п.) в текстах не встречаются (или, точнее, частотность их появления в текстах ниже порога отмеченности h), а конечный нуллолитерон их словарной трансформации в текстах либо — редко — обозначает паузу, либо — гораздо чаще — не имеет никакого соответствия в речи. Такие полилитеремы будем называть псевдолитеремами. Псевдолитеремой (псевдотрилитеремой) является, например, русский и болгарский предлог $][в][$. Трилитереме $][в][$ нельзя включить в орфографический словарь в качестве трансформы (репрезентанта), например, трилитерем $][в:$ или $][в?$. Аллолитер этих трилитерем в болгарских и русских текстах вообще не бывает. Снять орфографическую неопределенность псевдолитеремы $][в][$ удастся лишь путем перехода к орфотрафемам, соотнесенным с тетралитеремами или даже пенталитеремами типов $][в][м$ и $][в][сд$. Очевидно, предлог $в$ орфографические словари нельзя включать без специальных оговорок.

Все же для псевдолитеремы можно сформулировать орфографическое (точнее, анагностическое или анаграфическое) правило, по которому будут определяться некоторые инвариантные компоненты всех речевых соответствий псевдолитеремы. Такое орфографическое правило будем называть «парциальной (частичной) орфографемой» (соответственно «парциальной орфоанагностемой» или «парциальной орфоанаграфемой»). Например, для псевдолитеремы $][в][$ можно сформулировать парциальную орфоанагностему: «трилитереме $][в][$ соответствует в речи лабиодентал».

Полная формулировка орфографемы предполагает перечисление всех элементов множества Φ_j и всех элементов множества Γ_i , практически это нереализуемо. Поэтому роль орфографических словарей сводится к трансформации слов (репрезентантов полилитерем) официальной орфографии в полилитеремы орфографии более экономной, описывающейся при помощи меньшего числа правил (орфографем). Это словарная фонетическая транскрипция. Уменьшение числа орфоанаграфем в фонетической транскрипции достигается благодаря тому, что, как уже указывалось, многообразие всячески комбинируемых и принимаемых обществом в каче-

стве неслучайно отличных друг от друга звуковых фактов не очень большое — оно порядка 11 битов. А поскольку в обычных алфавитах количество букв порядка 30 (без учета перилитеронов, т. е. диакритических знаков), то построить не очень отличную от официальной и в то же время очень экономную орфографию — фонетическую транскрипцию — можно лишь на основе дилитерем и трилитерем, составленных из обычных 30 букв и некоторых перилитеронов. Звуковые соответствия литерем этой транскрипции вводятся в сознание людей в процессе обучения прямой речевой демонстрацией чтения и, поскольку число литерем фонетической транскрипции небольшое, просто запоминаются.

Нельзя ли заменить официальную орфографию «фонетической транскрипцией»? Последняя, правда, не столь экономна, сколь возможно теоретически, но зато не слишком отличается от официального правописания.

Наиболее серьезное препятствие на пути к принципиальному облегчению орфографии представляет собой фонология. Фонология возникла и развивалась по существу как теория орфографии, т. е. как теория трансформаций речевых сообщений в письменные и письменных сообщения в речевые. Однако, не имея возможности наблюдать непосредственно речевые сообщения, лингвисты до самого последнего времени исходили не из самих речевых сообщений, а из некоторых письменных буквенных трансформаций этих сообщений, называемых «точными записями речи». Таким образом буквенная дискретность негласно привносилась в объект лингвистики, проблема о собственной дискретности речи фактически подменялась проблемами, связанными с гомоморфностью, т. е. много-однозначностью трансформации «более скрупулезная буквенная запись речи — обычная буквенная запись речи». Речевое сообщение как результат изучаемых трансформаций по существу вообще исчезало из поля зрения лингвиста, поскольку результатом обеих трансформаций — приведенной и обратной ей — является письменное сообщение.

«Аспект чтеца» весьма важен для орфографии. Как уже было указано, число тех анаграфических правил (орфоанаграфем), при помощи которых в данной орфографии достигается изоанаграфия (безошибочное записывание), может быть весьма различным в разных орфографиях. Естественно, что легче изучить ту орфографию, которой при помощи меньшего числа правил достигается изоанаграфия. На первый взгляд может показаться, что это число является надежным критерием степени трудности орфографии, а следовательно, и степени возможного облегчения орфографии. В действительности же этот критерий может подвести, так как он

позволяет рассматривать как крупную орфографическую реформу устранение некоторого количества редко применяющихся анаграфических правил (орфоанаграфем). Да и практически невозможно быть вполне уверенным в том, что список орфографических правил, принятых во внимание исследователем, исчерпывающ для данной орфографии. Дело в том, что понятие «текст данного естественного человеческого языка» не является понятием с резко очерченным объемом. Всегда возможны споры относительно принадлежности к «данному языку» некоторых текстов слишком специальных наук или региональной литературы. Наконец, надо подчеркнуть, что на основе традиционной грамматической терминологии часто удается выразить большие множества орфоанаграфем одной и к тому же весьма компактной формулировкой, а это создает ложное впечатление «легкости» орфографии. Вот почему необходимо принимать во внимание частоту применения орфографических правил в среднем. Затем легко можно было бы находить вероятность появления анаграфической и/или анагностической неопределенности при устранении из орфографии какой-нибудь группы орфографем.

Если бы лингвисты при характеристике так называемых «фонетических орфографий» имели в виду не только правила записывания (анаграфические правила, орфоанаграфемы), но и анагностические правила, то не приводили бы в качестве «образца» фонетической орфографии сербскохорватское правописание. Ведь эта орфография не предусматривает обозначения так называемых «акцентов» и «долгот» гласных, вследствие чего вероятность появления анагностической неопределенности (средняя частота омографических речевых фактов) близка к 0,4 — из десяти букв (включая пробелы) сербскохорватского текста приблизительно четыре нельзя прочесть правильно на основании обзримого числа орфоанагностем. Подобным же образом обстоит дело с рядом других славянских орфографий — словенской, македонской, болгарской, русской, украинской, белорусской.

Более удачными с анагностической точки зрения (т. е. с точки зрения чтеца) являются чешская, словацкая и в несколько меньшей мере польская, а также лужицкие орфографии. С точки зрения чтеца более определенной оказывается даже известная своими традиционализмами французская орфография, а испанская орфография может служить примером реализации орфографии с малой степенью анагностической неопределенности.

Если абстрагироваться от псевдопроблемы, являются ли так называемые ударные гласные реализациями фонем, отличных от соответствующих им «безударных» фонем, обозначаемых на пись-

ме одними и теми же буквами (само понятие «фонема» является порождением особенностей орфографии и поэтому орфография может и должна анализироваться без помощи понятия фонемы), теоретический вопрос о причинах игнорирования интересов чтеца в новых кириллических славянских орфографиях может получить простой конкретно-исторический ответ: все они развились из русской «гражданской» азбуки, образцом которой служила графика без ударений¹⁴.

Если, опираясь на строгую теорию орфографий, разграничивать правила записывания и правила чтения, то, возможно, удастся построить и ввести в практику две орфографии: 1) более легкую для записи — предполагающую читателя, для которого соответствующий звуковой язык является родным; 2) более трудную для пишущего — дающую возможность читателям, для которых соответствующий язык не является родным, легко получать от письменного текста максимальную информацию о его звучании.

14 В своей статье «Об усовершенствовании русской орфографии» (ВЯ, 1963, 2, стр. 85) М. В. Панов, признавая за ударением достоинство «фонематической единицы», отмечает, что в «строгo последовательном фонематическом письме оно должно быть обозначено». Однако М. В. Панов не предлагает ввести в русскую орфографию обозначение словесного ударения. Это происходит из-за невнимания к анагностической стороне орфографии.