МЕТОДЫ ОБРАБОТКИ ИНФОРМАЦИИ
(Вычислительные системы)
1978 год Выпуск 74

УДК 655.25+801.4

АВТОМАТИЧЕСКОЕ ОПРЕДЕЛЕНИЕ МЕСТА ПЕРЕНОСА В ПРЕДЛОЖЕНИИ

Л.С.Юдина, А.С.Нудельман

В связи с дальнейшим увеличением потока информации в последнее время значительно возрос интерес специалистов к автоматизации редакционно-издательской деятельности (см.,например, [1,2]).

Предлагаемая работа также выполнена в плане проблемы совершенствования редакционно-издательской деятельности на базе современной вычислительной техники и является составной частью проекта ПАРИС [3].

Проект ПАРИС (Полный Автоматизированный Редакционно-Издательский Сервис) представляет собой попытку разработать общие принципы и конкретные пути использования ЭВМ для ускорения и повышения качества редакционно-издательской подготовки рукописи к печати ротапринтным, фотонаборным и другими способами [3].

В данной статье предлагается алгоритм автоматического принятия решения о переносе на другую строку части слова в соответствии с правилами русской орфографии. Потребность в этом возникает в процессе печати при автоматизированном выравнивании строки.

Из известных авторам работ, предшествующих данной, следует назвать [1]. Отличие их - прежде всего в назначении: алгоритм, описанный в [1], разработан для систем программирования набора, данный алгоритм ориентирован на ротапринтный способ издания. Кроме того, алгоритм [1] определяет места переноса только внутри слова, предлагаемый в данной работе - как в слове, так и в предложении в целом. Можно назвать также некоторые частности, которые в конечном счете отражаются на результатах работы: по-разному организован массив приставок и метод проверки на него; различаются наборы формализуемых правил, составляющих основу алгоритма.

Основное содержание работы составляет обсуждение общих принципов переноса и возможностей их алгоритмизации, а также описание самого алгоритма и экспериментов по автоматизации переносов в реальном тексте.

I. Общие принципы переносов

Задача автоматического определения места переноса должна включать два случая:

а) перенос внутри слова;

б) перенос между словами.

Авторами разработаны алгоритм и программа, осуществляющие автоматическое определение места переноса как внутри слова, так и между словами. Алгоритм реализован на ЭВМ "Минск-32". Следует отметить, что при написании программы преследовалась основная цель - добиться высокого качества переносов. По времени работы программа не оптимизировалась.

I.I. П е р е н о с в н у т р и с л о в а. Алгоритм автоматизации переносов слов основан на формализации существующих правил переносов [4,5]. Известно, что в основе действующих в настоящее время правил переносов лежат как фонетический, так и морфологический принципы [6]. Перенос внутри слова осуществляется, главным образом, по слогам (фонетический принцип). Однако это правило ограничивается запрещением переносить слоги, если они состоят из одного звука (буквы), и разрывать морфемы (морфологический принцип).

История развития языка свидетельствует о том, что стихийно появившиеся переносы усложнялись постепенно: ранние писцы при необходимости переносили любую оставшуюся часть текста. Впоследствии стали переносить по слогам, еще позже - с учетом морфемного состава слов. Реформа 1917-1918 гг. свела правила переносов к переносам по слогам (с небольшими ограничениями). Однако опять-таки стихийно стали появляться и поморфемные переносы. Считается, что существующие к настоящему времени правила переносов учитывают оба этих момента - слогоделение и морфемный состав слов [6]. Дело осложняется еще тем, что толкование самого слогоделения не единообразно.

В результате издательская практика содержит многочисленные отступления от правил в виде повсеместно встречающихся переносов

типа: нап-менования, ох-ранять, сог-ласования, ук-репление, на-более, оп-равдана, энергос-троя. Подтверждающаяся практикой нестрогость действующих правил переносов усложняет их формализацию.

Следует отметить, что авторы старались максимально придерживаться существующих правил переносов; пересмотр их в задачи работы не в коей мере не входил.

К настоящему времени считается установленным, что наименьшими произносительными единицами, на которые распадается речь, являются слоги, и слова следует переносить по слогам. Известно также, что наиболее распространенный, "классический" тип русского слога - сочетание согласного (С) с гласным (Г) - (СГ).

Возьмем эти предпосылки за основу алгоритмизации. Формально, без учета морфемной структуры, слово можно рассматривать как цепочку из элементов С и Г в различных (но фиксированных) комбинациях.

В задаче формализации переносов, очевидно, было бы оптимальным оперировать трех- и четырехбуквенными сочетаниями, однако проблемы сочетаемости в русском языке изучены ещё недостаточно, и мы вынуждены ограничиться биграммами с привлечением по мере надобности информации о предшествующих и последующих буквах.

На уровне биграмм сочетания внутри слова представлены типами: ГГ, СГ, ГС, СС, ␣Г, ␣С, Г␣, С␣, где символ ␣ обозначает пробел - начало или конец слова. Относительно сочетаний, содержащих его, принимается решение на основании правила: одна буква не оставляется на строке и не переносится, поскольку она ассоциируется в первую очередь не со словом, а с союзом, предлогом, междометием.

В числе оставшихся сочетаний группа ГГ самая малочисленная; сочетания ГГ для русского языка не характерны, они встречаются, главным образом, в заимствованных словах. В сочетаниях гласных алгоритмом предусмотрен перенос как между ними, так и после последней гласной, за исключением двух случаев, когда после последней гласной идет сочетание: а) согласный плюс (ь, ъ); два согласных, но не -ск-, -ст-, -гр-.

В сочетаниях СГ согласный всегда отходит к последующему гласному, и перенос между ними запрещен. Если при этом иметь в виду, что более половины общего числа русских слогов*) составляют слоги

*) Данные получены для русской речи, но с известными допущениями их можно распространить на печатный текст.

типа СГ [7], то, очевидно, можно заведомо ожидать свыше 50% правильных переносов.

Сочетания ГС делятся переносом, кроме случаев, когда после согласного идет согласный или (ъ, ь, й).

В сочетаниях СС – тоже немногочисленной группе сочетаний в русском языке – перенос предусмотрен между ними. Следовательно, в сочетаниях из трех и более согласных перенос может быть поставлен после каждой из них, кроме последней.

По-видимому, четкую грань между фонетическим и морфологическим принципами провести нельзя (это, собственно, отражено в самой истории вопроса и в формулировках правил переносов). При разработке алгоритма такие имели место случаи, требующие обоих подходов одновременно. Так, в алгоритме не делятся переносом сочетания согласных -от-, -ск-, -гр- – наиболее частые двухбуквенные сочетания согласных в русском языке [7].

Есть основания полагать, что они, как правило, входят в состав одних морфем; пример тому – -гр-, -ск-, -ст- в наиболее продуктивных морфемах (во вторых основах сложных слов): -граф, -грамм, -скоп, -стат, а также -ск-, -ст- в качестве одних из самых продуктивных словообразовательных суффиксов.

Кроме описанного, в алгоритме учтены такие следующие правила:

а) не оставляется на строке и не переносится часть слова, состоящая из одних согласных;

б) буквы (ъ, ь, й) не отделяются от предшествующей буквы;

в) две одинаковые согласные делятся переносом, за исключением двойных согласных корня, идущих после приставки;

г) часть слова, начинающаяся с (й), не переносится.

Наибольшие трудности вызывает, естественно, реализация морфологического принципа. Конкретно это вызвано необходимостью учитывать морфемную структуру слова, а именно:

– не разбивать переносом односложную приставку;

– не отделять от корня начальную часть, не составляющую слога;

– в сложных словах не отделять переносом начальную часть второй основы, не составляющую слога;

– не разбивать переносом односложную часть сложносокращенного слова.

Это и определяет порядок работы алгоритма: в память машины был введен "словарь приставок" – наиболее употребительные приста-

ки и наиболее частые элементы сложных слов [9]. Назовем их все квазиприставками (перечень их приводится в приложении I). Введение в память машины этого более или менее обширного "списка приставок" соответственно снижает процент ошибок, т.е. повышает качество машинных переносов.

Сначала слово анализируется на наличие квазиприставки, после которой ставится перенос; затем описанным выше способом происходит деление на слоги как квазиприставки, если в ней более одного слога, так и оставшейся части слова.

В словах с приставками, не помещенными в память машины, перенос осуществляется по описанным выше фонетическим правилам.

Требования морфологии при этом могут быть либо учтены (автоматически), либо не выполнены.

К случаям, не поддающимся формализации по общим принципам и являющим собой примеры неправильных "машинных" переносов, относится прежде всего следующие:

- сложные слова, поскольку все разнообразие частных, иногда единичных, случаев написания ввести в память машины невозможно, например: пищев-кусовой, пенкос-ниматель, солимот-рис *); в их числе следует особо отметить заимствованные слова и термины, составляющие, как известно, в русском языке особую языковую подсистему [10], например, ай-о-берг, бран-д-мауэр, фенол-ф-талеин;

- одинаковая реализация на письме неодинаковых по морфемному составу частей слова, для различения которых простых формальных признаков найти пока не удается: двухвостка, но двухвершинный, неофициальный - неологизм; не поддается такое формальному разделению случай, когда одно и то же буквосочетание в одних словах является приставкой, в других входит в состав корня: регресс - резной; в последнем случае на наиболее частые буквосочетания, в х о д я щ и е в с о с т а в к о р н я, производится специальная проверка с целью "отмены действия приставки", и слово переносится по общему принципу: сол-и-ще (но со-ждать).

I.2. П е р е н о с м е ж д у с л о в а м и. Как известно [4,8], не любое слово и не все, что написано через пробел или де-

---

*) Здесь указаны места только неправильных переносов.

фис, может быть разделено переносом. Наиболее распространенными примерами этого положения являются:

I. Аббревиатуры, написанные:

а) прописными буквами (ЦК ВЛКСМ),

б) частью прописными, частью строчными (КЗоТ),

в) прописными с цифрами, дефисом (ТУ-I04).

2. Наращения – грамматические окончания, соединенные дефисом с цифрами (2-й).

3. Единицы измерения – сокращенные обозначения мер, пишущиеся вместе с цифрами, указывающими число измеряемых единиц (I0 кг, 20 см).

4. Общепринятые графические сокращения (т.е., с-во, и т.п.).

5. Сокращенные слова в сочетаниях с именами собственными (г.Пушкин, пл.Свердлова).

6. Цифры или буквы со скобкой или точкой (при перечислении) в сочетании со следующими за ними словами (3. При наборе..., б) сокращенные выражения...).

7. Знаки и обозначения (§, №, %, С, ^{0}) в сочетаниях с цифрами.

8. Инициалы, фамилия и инициалы.

9. Пунктуационные знаки.

I0. Открывающие скобки и открывающие кавычки, а также знак сноски.

Подробнее формальные правила переносов между словами (через пробел) приведены при описании соответствующего алгоритма (п.2).

2. Алгоритм автоматического анализа текста

2.I. И с х о д н ы е о п р е д е л е н и я. Алфавит – полный набор различных символов, из которых составлен текст. В нашем случае это: строчные и прописные буквы русского алфавита (кроме "ё" и "Ё"); арабские цифры; знаки препинания (. , : ! ?); скобки ([, (,) ,]); кавычки ("); знаки параграфа, номера, процента (§, №,%); наклонная черта (/); минус (-); пробел (␣); пробел индексный (␣^{И}); пробел формульный (␣^{Ф}).

Формальный текст (или просто текст) есть последовательность символов алфавита, составленная в соответствии с общеупотребительными правилами машинописи. Все особенности оформления входного текста перечислены ниже:

1) Не допускается печать слов вразрядку (разрядку можно обозначать кодом).

2) Наклонная черта не используется в качестве скобки.

3) Пробел эквивалентен по употреблению "пропуску символа" в машинописном тексте.

4) Пробел индексный ставится вместо любого символа, приподнятого над строкой (опущенного под строку) на один интервал в машинописном тексте.

5) Пробел формульный резервирует место в строке, равное ширине символа, для вписывания формулы.

6) Минус, стоящий между пробелами, обозначает математический символ во вписываемой формуле, на котором возможен перенос.

7) Минус не используется в качестве знака переноса.

Отсутствие в формальном тексте знака переноса говорит о том, что формальный текст всегда выполняется в виде одной строки.

ПРИМЕР. Машинописный текст:

"Среди вершин[жж] множества $\Gamma(x_1)$ выби-
раем одну из вершин x, для которой
$\lambda'(x_1) - \lambda'(x) = \lambda(x,x_1) + (21 + 1)$,
и обозначаем выбранную вершину через x_2''

в формальном варианте выглядит так:

"Среди␣вершин ␣множества␣Ф Ф Ф Ф␣
выбираем␣одну␣из␣вершин␣Ф, ␣для␣которой
␣Ф Ф Ф Ф - Ф Ф Ф Ф - Ф Ф Ф Ф Ф - Ф Ф Ф
Ф Ф Ф, ␣и␣обозначаем␣выбранную␣вершину␣
через␣Ф Ф".

Любая буква русского алфавита является либо гласной, либо согласной, либо особой (й, ъ, ь).

Априори выбранную последовательность букв, являющуюся началом некоторого слова, будем называть квазиприставкой. Каждая квазиприставка либо является приставкой, либо не является таковой. Например, из выбранных в настоящей работе квазиприставок - "под" и "за" - приставки, а "низве" и "авиа" - не приставки.

Каждой квазиприставке Q априори ставится в соответствие множество Z(Q) запрещающих (выделение этой квазиприставки) буквосочетаний (это множество может быть и пустым; все непустые множества запрещающих буквосочетаний описаны в приложении I). В настоящей работе выбраны, например, $Z(\text{под}) = \emptyset$, $Z(\text{за}) = \{\text{пч}, \text{рп}\}$, Z (низве) =

= {ня}, Z(авня) = ∅ . Отсюда следует, что последовательность букв "за" в слове "забет" является квазиприставкой, а в слове "зарплата" не является таковой.

В работе тоже авторами фиксированы следующие множества:

а) множество нерасчленяемых буквосочетаний U = {гр, ск, от};

б) множество сокращений первого рода

S1 = {г., гор., гр., им., о., пос., с., т., тт., тов.};

в) множество сокращений второго рода S2 = {др., пр., т.д., т.п.}.

В дальнейшем символы формального текста будем обозначать через индексированные a, b и c. Предполагается, что символы любого текста пронумерованы в естественном порядке (номер первого символа больше или равен I). Условимся номер символа в тексте отождествлять с индексом. Таким образом, любой текст T длины n можно представить как последовательность $T = a_i a_{i+1} \dots a_{i+n-1}$.

Текст, составленный только из букв, будем называть словом.

Пусть $T = a_1 a_2 \dots a_n$ – текст. Символ a_i будем называть переносимым в данном тексте, если часть текста $a_i a_{i+1} \dots a_n$ можно перенести на следующую строку. Если такого переноса сделать нельзя, то a_i считается непереносимым.

2.2. Алгоритм выделения квазиприставки.

Входная информация:

а) $W = a_1 a_2 \dots a_n$ – слово длины n (n > 0);

б) $Q_1, Q_2, \dots, Q_m$ – перечень всех квазиприставок такой, что $q_{i+1} \le q_i$, где q_i – длина квазиприставки Q_i;

в) $Z(Q_i)$, $1 \le i \le m$, – множества запрещающих буквосочетаний.

Выходная информация: целое число q (q ≥ 0), при этом q > 0 означает, что последовательность $a_1 a_2 \dots a_q$ является квазиприставкой максимальной длины в слове W; q = 0 означает, что слово W не содержит квазиприставки.

Общая идея алгоритма: перебором в множестве квазиприставок ведется поиск квазиприставки (максимальной длины), совпадающей с началом слова W и такой, что в слове W после этой квазиприставки нет запрещающего ее буквосочетания.

Работа алгоритма состоит в следующем:

Шаг I. i := I.

Шаг 2 Если i > m, то q := 0 и шаг II.

Шаг 3. Если $q_i > n$, то шаг 10.

Шаг 4. Пусть $Q_i = b_1 b_2 \ldots b_k$, тогда если $\forall j (1 \le j \le k \to b_j = a_j)$, то шаг 5, иначе шаг 10.

Шаг 5. Если в $Z(Q_i)$ имеется буквосочетание $c_1 c_2 \ldots c_p$ такое, что $\forall j (1 \le j \le p \to c_j = a_{k+j})$, то шаг 10.

Шаг 6. Если $q_i > n - 3$, то $q := 0$ и шаг 11.

Шаг 7. Если среди букв $a_{q_i+1}, a_{q_i+2}, \ldots, a_n$ нет ни одной гласной, то $q := 0$ и шаг 11.

Шаг 8. Если a_{q_i+1} – особая буква, то $q := q_i + 1$ и шаг 11.

Шаг 9. Если Q_i – приставка, a_{q_i} – согласная и a_{q_i+1} – гласная буква, то $q := q_i - 1$ и шаг 11, иначе $q := q_i$ и шаг 11.

Шаг 10. $i := i+1$ и шаг 2.

Шаг 11. Конец.

2.3. А л г о р и т м а н а л и з а б у к в ы.

Входная информация:

а) $W = a_1 a_2 \ldots a_n$ – слово длины n ($n \ge 4$);

б) U – множество нерасчленяемых буквосочетаний;

в) m – номер символа в W, $3 \le m \le n$.

Выходная информация: целое число q ($q = 0$ или I), при этом $q = 0$ означает, что буква a_m непереносима в слове W; $q = I$ означает, что a_m переносима в W.

Общая идея алгоритма: решение о переносимости или непереносимости буквы a_m в слове W принимается на основе анализа части слова $a_{m-2} a_{m-1} a_m a_{m+1}$ в терминах гласная, согласная и особая (буквы), при этом учитывается расположение в этой части слова нерасчленяемого буквосочетания (если такое есть).

Работа алгоритма заключается в следующем:

Шаг I. Если a_m – особая буква, то $q := 0$ и шаг 8.

Шаг 2. Если среди $a_1, a_2, \ldots, a_{m-1}$ нет гласных или среди $a_m, a_{m+1}, \ldots, a_n$ нет гласных, то $q := 0$ и шаг 8.

Шаг 3. Если a_m – гласная, тогда, если a_{m-1} – согласная, то $q := 0$ и шаг 8, иначе $q := I$ и шаг 8.

Шаг 4. Если $a_{m+1} = a_m$, то $q := 0$ и шаг 8.

Шаг 5. Если a_{m-1} – особая буква, то $q := I$ и шаг 8.

Шаг 6. Если a_{m-1} – гласная, тогда, если a_{m+1} – особая или a_{m+1} – согласная и буквосочетание $a_m a_{m+1}$ не содержится в U, то $q := 0$ и шаг 8, иначе $q := I$ и шаг 8.

Шаг 7. Если $a_{n-2} = a_{n-1}$ или буквосочетание $a_{n-1}a_n$ содержится в U, то q:= 0 и шаг 8, иначе q:= 1 и шаг 8.

Шаг 8. Конец.

2.4. Алгоритм анализа слова.

Входная информация:

а) $T = a_1a_2 \dots a_n$ – текст длины n ($n \geq 4$);

б) $W = a_k a_{k+1} \dots a_{k+m-1}$ – слово длины m в тексте T ($m \geq 4$);

в) q_0 – номер символа в тексте T;

г) АВК (W', q) – алгоритм выделения квазиприставки (W'– входное слово, q – выходное число);

д) ААБ (W',i, q) – алгоритм анализа буквы (W' – входное слово, i – номер символа в W', q – выходное число).

Выходная информация: целые числа q^- и q^+ ($q^- \geq 0$, $q^+ \leq n+1$), при этом $q^- = 0$ ($q^+ = n+1$) означает, что не существует номера i такого, что $k < i \leq k+m-1$, $i \leq q_0$ ($i \geq q_0$), и a_1 переносим в слове W (и, следовательно, в тексте T); $q^- > 0$ и $q^- \neq q_0$ ($q^+ < n+1$ и $q^+ \neq q_0$) означает, что символ a_{q^-} (a_{q^+}) из слова $a_{k+1}a_{k+2} \dots a_{k+m-1}$ является ближайшим слева (справа) к a_{q_0} символом, переносимым в W (в T).

Если $k < q_0 \leq k+m-1$ и символ a_{q_0} из слова W переносим в W, то $q^- = q^+ = q_0$.

Общая идея алгоритма: слово W рассматривается как состоящее из квазиприставок и остатка, т.е. $W = K_1K_2 \dots K_1 R$, где K_1 – квазиприставка в W, K_2 – квазиприставка в W без K_1 и т.д., а R – часть слова W, в которой отсутствует квазиприставка; буква, стоящая после неоднобуквенной квазиприставки, переносима в W, вторая и последняя буквы в K_1 (и R) непереносимы в W, переносимость в W всех остальных букв слова W совпадает с переносимостью этих букв в соответствующей квазиприставке K_1 (остатке R).

Будем обозначать через MI множество номеров таких, что если $i \in$ MI, то символ a_i переносим в W (в T).

Работа алгоритма состоит в следующем:

Шаг 1. t:= k и MI :=$\emptyset$.

Шаг 2. Слово $W_1 := b_1b_2 \dots b_r$ ($b_i = a_{t+i-1}$, $r = m+k-t$).

Шаг 3. АВК (W_1, p).

Шаг 4. Если p = 0, то шаг 9.

Шаг 5. Слово $W_2 := b_1b_2 \dots b_p$.

Шаг 6. Для всех j, $3 \leq j < p$, ААБ (W_2,j,h), и если h = 1, то MI:=MI $\cup$ { t + j – 1 }.

Шаг 7. Если $p > 1$, то МП:=МП $\cup \{ t + p \}$.

Шаг 8. $t := t+p$ и шаг 2.

Шаг 9. Для всех j, $3 \le j \le r$, ААБ (a_1, j, h), и если $h = I$, то МП:=МП$\cup \{ t + j - 1 \}$.

Шаг IO. Из анализа множества МП устанавливаются значения q^- и q^+.

Шаг II. Конец.

2.5. А л г о р и т м а н а л и з а т е к с т а.

Входная информация:

а) $T = a_1 a_2 \ldots a_n$ - текст длины n ($n \ge 6$ и символы a_1 и a_n - пробелы);

б) q_0 - номер символа в тексте T;

в) ААС (W', q) - алгоритм анализа слова (W' - входное слово, q - входной номер).

Выходная информация: целые числа q^- и q^+ ($q^- \ge 0$, $q^+ \le n+1$), при этом $q^- = 0$ ($q^+ = n+1$) означает, что не существует номера i такого, что $4 < i < n$, $i \le q_0$ ($i \ge q_0$), и a_i переносим в тексте T; $q^- > 0$ и $q^- \ne q_0$ ($q^+ < n+1$ и $q^+ \ne q_0$) означает, что символ a_{q^-}(a_{q^+}) из текста T является ближайшим слева (справа) к a_{q_0} символом, переносимым в T.

Если $4 < q_0 < n$ и символ a_{q_0} переносим в T, то $q^- = q^+ = q_0$.

Общая идея алгоритма: могут быть переносимы в T только:

а) внутренняя буква некоторого слова из T, не являющегося единицей измерения при числе или аббревиатурой,

б) символ, стоящий после минуса,

в) символ, стоящий после пробела.

В случае "а" решение о переносимости принимает алгоритм анализа слова; случаи "б" и "в" рассматриваются данным алгоритмом непосредственно, при этом в обоих случаях анализируется контекст, содержащий исследуемый символ.

Через МП будем обозначать множество номеров таких, что если $i \in$ МП, то символ a_i возможен для переноса.

Слово $W = a_i a_{i+1} \ldots a_{i+m-1}$ из текста T будем называть максимальным, если в T a_{i-1} не буква и a_{i+m} не буква.

Работа алгоритма состоит в следующем:

Шаг I. МП:= $\emptyset$, $i := 4$.

Шаг 2. Если $i = n-1$, то шаг 28.

Шаг 3. Если a_i - буква, то шаг 8.

Шаг 4. Если a_i - минус, то шаг I0.

Шаг 5. Если a_i - пробел, то шаг I4.

Шаг 6. i:=i+1 и шаг 2.

Шаг 7. MI:=MIU{ i+1 } и шаг 6.

Шаг 8. Если a_{i-1} - буква, то шаг 6.

Шаг 9. Пусть $W = a_i a_{i+1} \ldots a_{i+m-1}$ - максимальное слово, тогда если $m \geq 4$, количество гласных в W больше единицы и среди a_{i+1}, $a_{i+2}, \ldots, a_{i+m-1}$ нет ни одной прописной буквы, то AAC (W, q_0), иначе шаг 6.

Шаг I0. Если a_{i-1} и a_{i+1} - цифры, либо a_{i-1} и a_{i+1} - пробелы формульные, то шаг 7.

Шаг II. Если a_{i-1} не буква или a_{i+1} не буква, то шаг 6.

Шаг I2. Пусть W_1 - ближайшее слева от a_i максимальное слово, а W_2 - ближайшее справа от a_i максимальное слово. Если W_1 или W_2 содержат только одну букву, или W_1 или W_2 не имеют гласных, или первая буква W_2 прописная, то шаг 6.

Шаг I3. Если W_1 содержит прописную не первую букву и W_2 содержит прописную не первую букву, то шаг 6, иначе шаг 7.

Шаг I4. Если a_{i+1} - пробел или %, или a_{i-1} - пробел или §, или №, то шаг 6.

Шаг I5. Если a_{i+1} - минус, тогда, если a_{i+2} - цифра, то шаг 7; если a_{i+2} - пробел, то шаг 6.

Шаг I6. Если a_{i-1} - скобка и a_{i-3} - пробел, то шаг 6.

Шаг I7. Если a_{i-1} не цифра и не пробел формульный, то шаг 22.

Шаг I8. Если a_{i+1} не пробел формульный, то шаг 20.

Шаг I9. Пусть $W' = a_{i+1} a_{i+2} \ldots a_{i+m-1}$ - последовательность формульных пробелов такая, что a_{i+m} не формульный пробел. Тогда, если a_{i+m} - минус, то шаг 7, иначе, если $m > 3$, то шаг 7, иначе шаг 6.

Шаг 20. Если a_{i+1} не буква, то шаг 7.

Шаг 2I. Пусть W_2 - ближайшее справа от a_i максимальное слово, тогда если в W_2 не более одной гласной, то шаг 6, иначе шаг 7.

Шаг 22. Если в тексте T слева от a_i нет ни одной буквы, то шаг 7.

Шаг 23. Если a_{i+1} не буква, то шаг 7.

Шаг 24. Если существует такое m, что последовательность $a_{i-m} a_{i-m+1} \ldots a_{i-1}$ есть сокращение первого рода и a_{i+1} - прописная буква, то шаг 6.

Шаг 25. Если существует такое m, что последовательность $a_{i+1}a_{i+2}\ldots a_{i+m}$ есть сокращение второго рода, то шаг 6.

Шаг 26. Если a_{i-1} не буква, то шаг 7.

Шаг 27. Пусть W_1 – ближайшее слева от a_i максимальное слово, а W_2 – ближайшее справа от a_i максимальное слово. Если W_1 и W_2 начинаются с прописных букв и W_2 состоит только из одной буквы, то шаг 6, иначе шаг 13.

Шаг 28. Из анализа множества МП устанавливаются значения q^- и q^+.

Шаг 29. Конец.

3. Эксперименты и их результаты

Алгоритм анализа текста был реализован в виде программы ЮСТРФ, оформленной на языке ФОРТРАН для ЭВМ "Минск-32". Эта программа предназначена, в частности, для использования в надпрограмме, формирующей строки фиксированной длины, и определяет в тексте места возможных переносов.

С целью проверки качества переносов, реализуемых программой ЮСТРФ, было проведено два эксперимента.

3.1. Эксперимент "слово". Цель эксперимента состояла в выяснении точности определяемых программой ЮСТРФ переносов внутри слов. Экспериментальный текст (3000 символов) был составлен из слов русского языка, выбранных случайным образом из словаря [11]. Экспериментальный текст представлен в приложении 2.

3.2. Эксперимент "текст". Цель эксперимента состояла в выяснении корректности определяемых программой ЮСТРФ переносов внутри произвольного (формального) текста. Экспериментальный текст (3000 символов) был взят из [12] со случайно выбранной страницы. Экспериментальный текст представлен в приложении 2.

3.3. Распечатка экспериментальных текстов. Из-за ограниченности набора символов УПч ЭВМ "Минск-32" при распечатке (формального) текста была использована следующая символика:

1. Прописная буква русского алфавита обозначает соответствующую строчную букву русского алфавита.

2. Прописная буква латинского алфавита обозначает аналогичную букву русского алфавита.

3. Вписанная буква обозначает прописную букву.

4. "Стрелка" – вопросительный знак.

5. "Равно" – пробел индексный.

6. "Больше" – прописную букву Б.

7. "Звездочка" – символ номера.

3.4. О ц е н к а р е з у л ь т а т о в э к с п е р и м е н т о в. При оценке качества "машинных" переносов производилось сравнение переносов, разрешенных программой ЮСТРФ, с переносами, допустимыми с точки зрения человека. Количественно такое сравнение выражается двумя числами: коэффициентом полноты и коэффициентом ошибки.

Пусть в тексте Т человек-специалист выделяет точно n переносимых (в Т) символов; пусть в этом же тексте программа (в частности, ЮСТРФ) относит к переносимым m символов, причём m_0 из них являются переносимыми и с точки зрения человека (понятно, что $m_0 \leq m$ и $m_0 \leq n$). Тогда коэффициент полноты

$$P = \frac{m_0}{n},$$

а коэффициент ошибки

$$R = \frac{m - m_0}{m}.$$

Следует отметить, что коэффициенты полноты и ошибки характеризуют различные стороны автоматического переноса и поэтому являются достаточно независимыми. Бо́льший коэффициент полноты указывает на бо́льшее использование "ресурсов" переноса, имеющихся в русской грамматике. При меньшем коэффициенте ошибки повышается грамотность разбивки текста на строки.

Результат эксперимента "слово": $P_1 = 0.963$, $R_1 = 0,053$ (при $n_1 = 750$).

Результат эксперимента "текст": $P_2 = 0,955$, $R_2 = 0,005$ (при $n_2 = 940$).

3.5. В ы в о д ы. Полнота и ошибочность переносов, вынимаемых программой ЮСТРФ, находится, по-видимому (см. I.I), в пределах требований существующих правил переносов. Для более надежной оценки качества машинных переносов необходимы экспериментальные тексты различного характера (технические, общественно-политические, художественные) объемом порядка 10^6 символов.

Анализ автоматических переносов, полученных в экспериментах, позволяет предположить, что их качество в основе своей удовлетворяет требованиям практики печати. В то же время преимущества автоматического определения места переноса в слове в свете проблемы программирования набора несомненны.

Авторы выражают благодарность Ю.Г.Косареву за постановку задачи.

Литература

I. БЕРЛИН А.С. Системы программирования набора. М., "Книга", 1971.

2. ВОРОЖЦОВА Т.П., ВОСКРЕСЕНСКИЙ М.И., КОСТИН Б.А. Автоматизированная система набора, правки и верстки. -"Полиграфия", 1973, № 2.

3. КОСАРЕВ Ю.Г. Проект ПАРИС - полный автоматизированный редакционно-издательский сервис. Отчет ИМ СО АН СССР, Новосибирск, 1974.

4. БЫЛИНСКИЙ К.И., НИКОЛЬСКИЙ Н.Н. Справочник по орфографии и пунктуации для работников печати. М., "Просвещение", 1970.

5. Правила русской орфографии и пунктуации. М., изд-во "Искусство", 1962.

6. ИВАНОВА В.Ф. Современный русский язык. Графика и орфография. М., "Учпедгиз", 1966.

7. ЦИНА Л.С. Автоматизация транскрипции и некоторые количественные характеристики русской речи. Автореф. дис. на соиск.учен. степени канд. филологич. наук. Л., 1974, 23 с. (Ленингр.гос.ун-т).

8. РОЗЕНТАЛЬ Д.Э. Справочник по правописанию и литературной правке. М., "Книга", 1971.

9. БУКЧИНА Б.З., КАЛАКУЦКАЯ Л.П. Сложные слова, М., "Наука", 1964.

10. ПАНОВ М.В. Русская фонетика. М., "Просвещение", 1967.

11. Орфографический словарь русского языка. М., изд-во "Советская энциклопедия", 1970.

12. Проблемы структурной лингвистики. М., изд-во АН СССР, 1963.

Поступила в ред.-изд.отд.
3 августа 1977 года

П Р И Л О Ж Е Н И Е I

Вспомогательные массивы, используемые для автоматизации переносов

```
МКЛ   001 001      КРИСТАЛЛО,ФОРМАЛЬНО.)
      001 002      РЕНТГЕНО;)
      001 003      ВОЗДУХО,РАЗЛОЖЕНО,ЧАСТИЧНО,МЕТАЛЛО,
      001 004      ОБРАТНО,ОСЦИЛЛО,ПЕРЕДНЕ,ПРОТИВО;
      001 005      СЕЛЬСКО,СМЕРАНО,УСЛОВНО,ЧЕТЫРЕХ,
      001 006      ЭЛЕКТРО;)
      001 007      БЫСТРО,РАДОНО,ВЗАИМНО,ВНУТРИ,ВУЗЛОВО,
      001 008      ГРЕЙФО,ДЛИННО,АРЕВНЕ,ЖЕЛЕЗО,ЗОЛОТО,
      001 009      КАРАНО,КЛЕТКО,КОНТРА,КРАСНО,КРЕПКО,
      001 010      КРУГЛО,КРУПНО,МАШИНО,НЕРВНО,ОРГАНО,
      001 011      ПЛОСКО,ПСЕВДО,ПРОЧНО,РЕЛЬСО,СИЛЬНО,
      001 012      СЛОЖНО,СРЕДНЕ,СТЕКЛО,СТЕРЕО,ТОЛСТО,
      001 013      ТРУДНО,ТЯЖЕЛО,УЛЬТРА,ЧАСТНО,ШИРОКО,
      001 014      ЭКСТРА,ЭНЕРГО,)
      001 015      БЛАГО,ВИБРО,ВЛАГО,РЕЛЬО,РЕКТО,ГИГРО,
      001 016      ГИДРО,ГИПЕР,ГОРНО,ГУСТО,ЕДИНО,ИЗЖИВ,
      001 017      ЗВАНЕ,ЗВУКО,ЗЕРНО,ЗЛАТО,КАЗНО,КАМНЕ,
      001 018      КВАЗИ,КОЖНО,КОНТР,КОСМО,КРОВО,КРУТО,
      001 019      ЛЕГКО,МАКРО,МЕЛКО,МИКРО,МНОГО,НЕФТЕ,
      001 020      НИЗКО,ОСТРО;ПЕРВО,ПЛОДО,ПРЯМО,ПУСТО,
      001 021      РАВНО,РАДИО,РЕЗКО,СВЕРЕ,СВЕТО,СВЕРХ,
      001 022      СЛАБО,СЛОВО,СТАРО,СТЕНО,СТЕТО,СТИХО,
      001 023      ТРОЛО,ТЕРМО;ТЕСНО,ТОНКО,ТРАНС,ТРУДО,
      001 024      ТУРБО,ХЛЕБО,ХРОМО,ХРОНО,ЦИКЛО,ЧЕРНО,
      001 025      ЧИСТО,ШЕЛКО,БЕНЗО,ГЕКСА,МЕТРО,ПРАВО,)
      001 026      АВИА,АВТО;АГРО,АНТИ;АРХИ;АЭРО;ВЕЛО,
      001 027      ВОДО,ГЛЫБЕ,ГАЗО,ГЕРО,ГИРО,ГОМО,АБУХ,
      001 028      ДИКО,ЖИВО,ЗУБО,КИНО,ЛЕВО,ЛЕСО,ЛИТО;
      001 029      ЛЬНО,МАЛО,МЕТА,МОНО,МОТО;МЯСО,
      001 030      НИРЕ,НОВО,НСЕО;ОБЩЕ,ОАНО,ОРТО;ОРФО,
      001 031      ПЕРЕ;ПИРО,ПОЛУ,ПРЕА,ПУТЕ,РУКО,САМО,
      001 032      ТЕЛЕ,ТРЕХ,УГЛЕ,УЗКО,ФИТО,ФОНО,ФСТО,
      001 033      ЦИТО,ЧУЖЕ,ЭНДО,)
      001 034      БИО,ВНЕ,ВСЕ,ГЕО,АВУ,ЕЖЕ,ЗЛО;ЗОО;ИЗО,ИНО,
      001 035      ЛИЕ,НАД,ОТО;ПОД,ПРЕ,ПРИ,ПРО,ЭКС,ЭПИ,)
      001 036      ВЫ;АЕ;АО;ЗА;НА,НЕ;ОЬ,ОТ,ПО;РЕ;СО;)
      001 037      0,)++++++++++++
      001 038      ++
МКШ   001 039      РЕНТГЕНО-ВС;
      001 040      ЭЛЕКТРО-НИ;
      001 041      ИЗЖИЕ-НИ;
      001 042      ОСТРО-ИИ;
      001 043      АВТО-РС;
      001 044      АНТИ-ИИ;
      001 045      АРХИ-ВИ;
      001 046      АЭРО-ВИ;
      001 047      ЛИТО-ВС;
      001 048      МОТО-РК,РИ;
      001 049      ПЕРЕ-АИ,ИИ;
      001 050      ЗЛО-СТ;
      001 051      ОТВ-ИА,ПЛ,ПТ;
      001 052      ВЫ-ИП;
      001 053      АЕ-ВС,ВЧ,ГТ,ЛЬ,НН,НТ,НЬ,РГ,РЖ,РЗ,РМ,РН,
      001 054      РР,СИ,СР,ТС;
      001 055      АО-БЛ,БР,ГМ,КТ,ЛБ,ЛГ,ЛЧ
      001 056      ,МР,НК,НИ,СК,ХЛ,ХИ,ЧК;
      001 057      ЗА-ПЧ,РП;
      001 058      НЕ-ВР,ВИ,КТ,МК,МЦ,МЧ,НЦ,РБ,РП;
      001 059      ПО-ЗИ,ЧИ;
      001 060      РЕ-АК,ЗЬ,ЛЬ,НТ;
      001 061      СО-БС,ВХ,ЛА,ЛИ,НЛ,НИ,РТ,ТИ,ХИ,ШИ;
      001 062      ++++++++++++++++++++++++++++
      001 063      ++
```

```
МЗ    001 064    ГР,СК,СТ;++++++++++++++++
      001 065    ++
МРЗ   001 066    1-39,2-18,3-00,4-00,
      001 067    Г.,ГОР.,РР.,ЫМ.,О.,ПОС.,С.,Т.,ТТ.,ТОВ-Т
      001 068    АР.,ПР.,Т.А.,Т.П.;
      001 069    +++++++++++++++++++++++++++++
      001 070    +++++++++++++++++++++++++++++
      001 071    ++++++++++++++
      001 072    ++
МН    001 073    0000010п2п03004005006007008009
      001 074    90+40-5п/10,11,70ш+0ш10+20(21)
      001 075    +0Е80-1п;20[21]30*-0щ-0щ-0Ч-0Х
      001 076    -0Б10:+1А-1Б-1В-1Г-1А+1Е-1Ж-13
      001 077    +1И/1Й-1К-1Л-1М-1Н+1О-1П-1Р-1С
      001 078    -1Т+1У-1Ф-1Х-1Ц-1Ч-1Ш-1Щ+1Ы/1Ь
      001 079    +1Э+1Ю+1Я-0А-0Б-0Г+0И/0Й-0Л-0Н
      001 080    +0П-0Р-0С+0У-0Ж-0В-03/1--0К+09
      001 081    -0Т/0Ь-0П+0О-0Ч-0Ы60%+0Ш22|+0А
      001 082    /0-10!++++++++++++++++++++++++
      001 083    ++++++++++++++++++++++++++++++
      001 084    ++++++++++++++++++++++++++++++
      001 085    ++++++++++++++++++++++++++++++
```

ПРИЛОЖЕНИЕ 2

Примеры автоматического нахождения допустимых мест переносов

Алгоритм установления связи слов в предложении

"5 чем связано это слово?" — такой вопрос нам приходилось слышать часто, а особенно часто от преподавателя латинского языка. Ответить на него было не просто, но первых, язык нам незнакомый, во-вторых, не ясно, как это связано. И этой главе предлагается процедура установления связи между словами, на основании которой мы сможем более ясно представить, что такое связь. Процедура включает операции, достаточно простые для того, чтобы их могла выполнить вычислительная машина, а процесс вычисления длинен. Поэтому осуществление этой процедуры по самой ее сути предназначено машине. И настоящем алгоритме допущения довольно сильны, но для получения многих требуемых сведений у нас уже имеются алгоритмы. И настоящем алгоритме предполагается известным деление слов на классы по классам R.L.Добрушина* и границы предложения (всегда имеется в виду простые предложения). При этом не надо знать, что класс 1 глагол, а класс 2 — наречие; достаточно знать что данное слово принадлежит к классу 1, а не 2. Мы считаем свои допущения приближенными, потому что у нас есть, хотя и не вполне совершенные, алгоритмы выделения морфем; кроме того, мы верим в то, что алгоритм выделения слов и алгоритм выделения предложения при известных морфемах создать не трудно. Поясним на примере, каких результатов мы ожидаем от алгоритма обнаружения связей слов. Пусть в нашем тексте встретилось предложение "Большой дом стоял на крутом берегу реки".

Примем следующие обозначения: глаголы — 1, именительный падеж существительного — 2, родительный падеж существительного — 3, предложный падеж существительного — 4, именительный падеж прилагательного — 5, предложный падеж прилагательного — 6, наречия — 7, предлоги — 8. Тогда наш пример можно будет записать так: 521786643. Нарисуем схему этого предложения, подобную той, которую нам показывали в школе: 1—2—5, 1—7, 1—8—9—3, 4—6. Черта между 5 и 2 показывает, что сочетание "большой дом" осмысленно; то, что нет черты между 5 и 8, показывает, что сочетание "большой на" не имеет смысла. Наш алгоритм должен найти именно такую схему для каждого предложения нашего текста, не пользуясь критерием осмысленности. Разумеется, схемы следует искать для предложений, содержащих более двух слов. Итак, для начала работы нам нужен текст, закодированный следующим образом: 1) каждому слову соответствует номер или символ данного класса, 2) между простыми предложениями стоит символ, обозначающий границу.

* R.L.Добрушин, "Математическая лингвистика", "Математическое просвещение", 1961, № 6. Однако классы, получаемые при такой классификации, очень мелки, и для экспериментов мы рекомендуем гораздо более крупные и менее правильные, например для русского языка: глагол в личной форме, каждый из падежей существительного, каждый из падежей прилагательного или причастия как отдельные классы, инфинитив, деепричастие, наречие, предлог, частица, союз (см. ниже).

ЭКСПЕРИМЕНТ "СЛОВО"

1. АБО+НЕН+Т+СКИЙ АВАН+ТЮ+РИ+СТ+СКИЙ АВИА+ЦИ+ОН+НЫЙ АВ+ТО+ПАРК АВ+Т
2. О+ЭЛЕК+Т+РОН+НЫЙ АДА+ПТИ+РО+ВАН+НЫЙ АД+РЕ+САНТ АЙ+С+БЕРГ АК+КОР+
3. Д+НЫЙ АК+СЕ+ЛЕ+РА+ТОР АЛЕТЬ АЛЬ+МА+НАХ АМА+РАН+ТО+ВЫЙ АМУ+НИ+ЦИ
4. Я АНА+НАС+НЫЙ АНЕК+ДО+ТИЧ+НОСТЬ АН+НЕК+СИ+РО+ВАТЬ+СЯ АН+ТИ+ПАТ+Р
5. И+О+ТИЧ+НЫЙ АН+Т+РА+ЦИТ+НЫЙ АП+ПА+РАТ АРА+БО+ЛОГ АР+КАН+ЗАС АР+
6. ТЕ+РИ+АЛЬ+НЫЙ АС+СИ+РИЙ+КА АУК+СИО+ПО+РА АЭ+РО+ПЛАН+НЫЙ БА+ЛАН+С
7. И+РОВ+КА БЕЗ+ГРА+НИЧ+НОСТЬ БЕ+ЗО+СТА+НО+ВОЧ+НЫЙ БЕС+ПО+ЩАД+НОСТЬ
8. БЕ+ТО+НОС+МЕ+СИ+ТЕЛЬ+НЫЙ БОЛ+ВАН+КА БРАН+Д+ДА+У+ЭР БРО+НЕ+ТИН Б
9. Ы+СТ+РО+СОХ+НУ+ЩИЙ ВА+ГО+НО+ОПРО+КИ+ДЫ+ВА+ТЕЛЬ ВВО+РА+ЧИ+ВАТЬ ВЕ+
10. СТ+ГО+ТЫ ВЗДЫБ+ЛИ+ВАТЬ+СЯ ВЗМО+СТИТЬ+СЯ ВКРАП+ЛЕН+НЫЙ ВО+ДО+СБОР
11. ВОЗ+Г+ЛАВ+ЛЯТЬ ВОП+ЛО+ЩЕ+НИЕ ВОС+П+РО+ИЗ+ВЕ+СТИ ВРАЗ+ЛЕТ ВСУ+Х
12. УЮ ВЫ+БУ+РАВ+ЛИ+ВАТЬ ВЫ+ЖЖЕН+НЫЙ ВЫ+ДОЛ+БЛЕН+НЫЙ ВЫ+ПЛА+КАТЬ+СЯ В
13. Ы+СО+КО+КВА+ЛИ+ФИ+ЦИ+РО+ВАН+НЫЙ ВЫ+ХОД+КА ГА+ЗО+ВЫ+СТ+КА ГИ+ПЕР+
14. ПЛА+ЗИЯ ГНУС+НЫЙ ГОР+НО+ПРО+ХОД+ЧЕ+СКИЙ ГРЕ+СТИ ГУ+САР+СКИЙ ДВУ
15. +ЛИЧ+НЫЙ ДЕ+КА+ДЕН+Т+СКИЙ ДЕР+МА+ТО+ЛО+ГОВ ДИ+ЗЕ+ЛИСТ ДОБ+РО+ВО+
16. ЛЕЦ ДОЛ+ТИ ДО+КЛЕ+ЕН+НЫЙ ДО+СТАТЬ+СЯ ДРЕ+НИ+РО+ВА+НИЕ ДУ+РА+ЛЕ+
17. НИ+НИ+Е+ВЫЙ ЖАТ+КА ЖИЗ+НЕН+НОСТЬ ЗА+ВАЛ+ЛАС+ТИ+РО+ВАТЬ ЗА+ГО+ВОР+
18. ЩИ+ЧЕ+СКИЙ ЗА+ЖИ+ГА+ТЕЛЬ+НЫЙ ЗА+КОН+СЕР+ВИ+РО+ВАТЬ ЗА+МЕД+ЛЯТЬ+СЯ
19. ЗА+ПА+САТЬ+СЯ ЗА+ПЛЕ+ТЕН+НЫЙ ЗА+СО+ЛКА ЗА+ТМЕ+ВАТЬ ЗА+ЧВА+НИТЬ+
20. СЯ ЗВЯ+КАТЬ ЗНА+МЕ+НИЕ ИГРОК ИЗ+Г+ЛА+ЖИ+ВАТЬ ИЗ+НУ+РЯТЬ ИМИ+ТИ
21. +РО+ВАТЬ ИН+СТ+РУК+ТИ+РО+ВАТЬ ИСК+РИ+СТЫЙ ИЗУ+МЛЕН+НЫЙ КА+ВА+ЧИ
22. Я КА+НАД+СКИЙ КАР+КАНЬЕ КА+ФЕЛЬ+ЩИ+ЦА КИ+НО+ФО+ТО+ДО+КУ+МЕ
23. НТ КЛИЧ+КА КОК+СО+ВЕН+ЗОЛ КО+МИ+ТЕТ+СКИЙ КОН+С+ПИ+РА+ТОР+СКИЙ К
24. ОНТР+РЕ+ВО+ЛЮ+ЦИ+ОН+НЕР КОР+РЕС+ПОН+ДИ+РУ+ЮЩИЙ КРАС+НО+ВА+ТЫЙ ЛИ+
25. ЛО+ВА+ТЫЙ КРИ+ЧА+ЩИЙ КУ+ВШИН КУ+ЛО+ЧЕК ЛА+МЕТ+НЫЙ ЛЕ+СО+ТУР+ГО+
26. ВЫЙ ЛИ+ТИЙ ЛО+СКУТ+НИ+КА МАГ+НИ+ТО+ЛИ+Е+ВЫЙ МЕ+ЛО+РОС+СИ+ЯН+КА МА+
27. СКИ+РОВ+ЩИК МЕД+САН+БАТ МЕР+ЛУ+ШЕЧ+НИК МИК+РО+ВИНТ МНО+ГО+ЗНА+ЧИ
28. +ТЕЛЬ+НОСТЬ МОЛЬ+БЕРТ МО+ТО+КРОСС МО+СО+ХРА+ДЕ+ФОН+БИ+НАТ НА+БРА
29. Н+НЫЙ НАД+МЕН+НОСТЬ НА+КАП+ЛИ+ВАТЬ+СЯ НА+МАС+ЛИ+ВАТЬ НА+ПО+ЯВАТЬ
30. +СЯ НА+САД+КА НА+СТУП+ЛЕ+НИЕ НА+ЦИ+О+НА+ЛИ+СТИЧ+НОСТЬ НЕ+ВНИ+МА+
31. НИЕ НЕ+ЗДО+РО+ВЫЙ НЕ+ОПЛА+ЧЕН+НЫЙ НЕ+СКОН+ЧА+Е+МЫЙ НИЗ+ВЕР+ГНУТ
32. Ь НО+МО+ГРАМ+МА ОБ+ВЯ+ЗЫ+ВАТЬ ОБ+ЖИТЬ ОБ+МУН+ДИ+РО+ВАТЬ ОБОН+НЕ
33. +ЛЬНЫЙ ОБ+СЕ+ВАТЬ ОБ+ЩЕ+УПОТ+РЕ+БИ+ТЕЛЬ+НОСТЬ ОГРЕХ ОБУ+ЧЕН+НЫЙ
34. ОКСИ+АН+МЕТ+РИЯ ОЛИ+ФОН+КА ОПИ+ЛЕН+НЫЙ ОСИ+ЛОСТЬ ОСТ+РО+СЛО+ВИТЬ
35. ОТ+ГО+ВО+РЕН+НЫЙ ОТ+КО+ПАН+НЫЙ ОТО+РВИ+ГОЛ ОТ+РЕ+ПЬЕ ОТ+ТО+ПЫ+РИ
36. ТЬ+СЯ О+ФАК+ТУ+РИТЬ ПА+ВИЛЬ+ОН+НЫЙ ПА+НУ+ЛЕ+СКИЙ ПАС+СА+ТНЫЙ П
37. ЕН+КОС+НИ+МА+ТЕЛЬ ПЕ+РЕ+ВОП+ЛО+ЩЕН+НЫЙ ПЕ+РЕ+ЗИ+МО+ВАН+НЫЙ ПЕ+РЕ+
38. ЛОМ+НЫЙ ПЕ+РЕ+НА+ЧО+ЛАН+НЫЙ ПЕ+РЕ+СОВЕТ ПЕ+РЕ+ЩИ+ПИ+ЦА ПЕ+ЧА+ТАН+НЫЙ
39. ПИ+НЕ+КУ+СО+ЧЕК ПЛО+СКОСТЬ ПО+БРЕЗ+ГАТЬ ПО+ГРЕБ ПОД+ЛА+ВОК П
40. О+ЛОЙ+ГУ ПОД+СЛЕТ ПОД+ТЯ+НУТЬ ПО+ЗНА+ВА+ЕМ+НЫЙ ПО+КРУ+ТИТЬ ПО+ЛИ
41. О+МО+ЧИЕ ПО+ЛУ+ФИ+НАЛ ПО+НА+ЧАЛЬ+СТ+ВО+ВАТЬ ПО+ПРИ+ГУН ПО+СЛЕ+ЧА
42. С ПО+СТУ+КАТЬ ПО+ХЛЕБ+КА ПРЕ+ДО+СУД+НЫЙ ПРЕД+СЕ+ДА+ТЕЛЬ ПРЕ+ССО
43. +ВЫЙ ПРИ+СКА+ЗАТЬ ПРИ+КЛЕ+ИВАТЬ+СЯ ПРИ+ДВО+РЕ+ЗВА+ТЕЛЬ ПРИ+СНИТЬ+СЯ
44. ПРИ+ТРА+ГИ+ВАТЬ+СЯ ПРО+БЛЕ+МА ПРО+ДАВ+ЛИ+ВАТЬ ПРО+ЛА+ЗА ПРО+ШТО+
45. ВА+РЫ ПРО+РВА+СИТЬ ПРО+СТЫ+НЯ ПРО+СТО+ФИ+ЛЯ ПРО+ТИ+ВО+ЛО+ДОЧ+НЫЙ
46. ПУ+СТЯК РА+ДЕТЬ РАЗ+ЛЕН+ЧАН+НЫЙ РАЗ+ГРА+ФИ+ТЬ РАС+СУ+ДИ+ТЕЛЬ+НО
47. СТЬ РАЗ+МНО+ЖИТЬ РА+ЗО+РИТЬ РА+СКАТ РАС+ПИ+СА+НИЕ РАС+СЛЕ+ДО+ВА
48. ТЬ РАС+Т+РЕ+ПАТЬ РЕ+ЗОН РЕ+СПЕК+ТА+БЕЛЬ+НОСТЬ РО+ГО+ЖА РУ+КО+ВО
49. +ДИ+ТЕЛЬ САВ+РА+СЫЙ СА+МО+РАЗ+ЛО+ЖЕ+НИЕ СВЯ+ЗАТЬ СВЕ+ТО+БО+ЯЗНЬ
50. СГРЕ+СТИ СЕ+МЕЙ+СТ+ВО СЕР+ТИ+ФИ+КАТ СИ+НЕ+ЛОМ+КОСТЬ СКО+ВЫ+ВА
51. ТЬ СКО+РО+ПЕ+ЧАТ+НЫЙ СКАН+ЦЕ+ПЕ+РЕ+ГОН+НЫЙ СМА+ЗО+ВА+НИЕ СМЯ+ТЕ+
52. НИЕ СО+ВНАР+ХОЗ СО+ЛО+МОТ+РЯС СО+ОТ+ВЕ+ДО+ТО+ЧЕН+НОСТЬ СПЕ+ЛОСТЬ
53. СПРО+ФИ+ЛИ+РО+ВАТЬ СТА+КАН+ЧИК СТЕ+КОЛЬ+ЩИК СТОЛЬ+КО СТРО+ЕНИЕ
54. СУ+ДО+ПОД+ЪЕМ СУ+ШЕ+НИ+КА СУ+РО+ВОЛ+КА ТА+РЕЛ+КО+ОБ+РАЗ+НЫЙ ТЕ+
55. О+ЛО+ГИЯ ТИ+ПО+ЛИ+ТО+ГРА+ФИЯ ТОН+ЧАЙ+ШИЙ ТРАНС+ПО+НИ+РО+ВА+НИЕ Т
56. РИ+КО+ТАЖ ТРУД+НО+ВО+ЛЬ+НОЙ УВА+ЖИ+ВАТЬ УГРО+РУС+СКИЙ УК+ЛА+ДОЧ
57. +КА УМИЛЬ+НЫЙ УПОР+СТ+ВО+ВАТЬ УС+МОТ+РЕТЬ УТОМ+ЛЯ+Е+МОСТЬ УШ+КУ
58. Й+НИ+ЧАТЬ ФЕ+НОЛ+ФОР+МАЛЬ+ДЕ+ГИД+НЫЙ ФЛЕК+СИ+РУ+Ю+ЩИЙ ФРАН+КМА+СОН+
59. СТ+ВО ХА+МИ+НЫЙ ЦЕ+РЕ+МО+НИЙ+МЕЙ+СТЕР ШКО+ЛЯР+СТ+ВО