

Как реализовать компьютерное "чувство юмора"?

И.М.Суслов

Институт физических проблем им. П. Л. Капицы РАН,

119334, Москва, Россия

E-mail: suslov@kapitza.ras.ru

Аннотация

Предложенная ранее компьютерная модель "чувства юмора"[1–3] доводится до уровня реального алгоритма.

1. Введение

В предыдущих работах автора [1, 2, 3] предложена схема обработки информации, естественным образом приводящая к возможности реализации в компьютерах простейших человеческих эмоций, и в частности — "чувства юмора". В настоящей работе мы попытаемся довести эту схему до уровня конкретного алгоритма.¹

Вкратце, сформулированная ранее модель [1] состоит в следующем. Пусть на вход процессора поступает последовательность символов (или "слов") $A_1, A_2, A_3 \dots$. С каждым словом A_n ассоциируется некоторый набор образов $\{B_n\}$. Задача состоит в выборе из каждого набора $\{B_n\}$ одного образа $B_n^{i_n}$, который подразумевается в данном контексте. Будем считать, что текст "понят", если последовательности слов $A_1, A_2, A_3 \dots$ сопоставлена последовательность образов $B_1^{i_1}, B_2^{i_2}, B_3^{i_3} \dots$, которую можно представлять как некоторую "траекторию"(рис. 1).

В принципиальном плане алгоритм состоит в следующем: (а) составляются всевозможные траектории; (б) каждой траектории приписывается некоторая вероятность; (в) выбирается траектория с максимальной вероятностью. Нетривиальным является лишь этап (б), т. е. алгоритм определения вероятности траектории. В основе такого алгоритма должна лежать информация о сочетаемости образов, которая приобретается в процессе "обучения", требующего введения в процессор достаточно длинного куска "расшифрованного" текста, т. е. текста, записанного не в словах, а в образах.

Любая конкретная реализация такого алгоритма требует числа операций, экспоненциально растущего с длиной текста; поэтому алгоритм позволяет обрабатывать

¹ В настоящей работе мы ограничимся простейшими образцами юмора, связанными с первичной обработкой информации. Высшие уровни обработки информации устроены аналогично, но требуют более сложных конструкций [3].

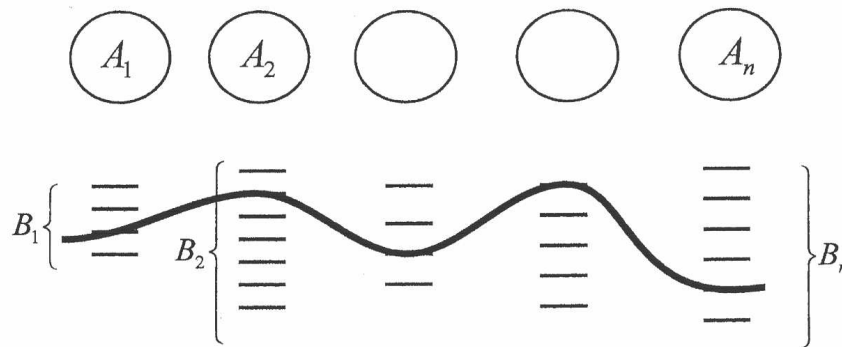


Рис. 1: Схема обработки информации: каждому из символов A_n сопоставляется набор образов $\{B_n\}$, из которого надо выбрать один образ $B_n^{i_n}$; последовательность $B_1^{i_1}, B_2^{i_2}, B_3^{i_3} \dots$ можно представлять как некоторую "траекторию".

отрезки текста, не превышающего некоторого числа L слов. Для обработки более длинных текстов естественно использовать следующую процедуру. При чтении первых L слов сохраняется не одна, а несколько (M) наиболее вероятных траекторий. Затем нужно сдвинуться на один шаг — рассмотреть текст от второго до $(L + 1)$ -го слова — и для каждой из M сохраненных траекторий построить всевозможные продолжения. Из них опять сохраняется M наиболее вероятных и т. д. В целом процесс выглядит так (рис. 2): вблизи переднего фронта A траектория сильно ветвится; в некоторой точке B ветвление заканчивается; наконец, расшифрованная траектория CD (с некоторой задержкой AC) поступает на выход процессора.

На первый взгляд, точка C должна всегда находиться позади точки B или совпадать с ней. Однако в реальных условиях функционирования биологических систем задержка AC должна быть ограничена сверху: обработка информации происходит в подсознании и до того, как в сознание выведена расшифрованная траектория CD , в него не поступает вообще никакой информации. Поэтому если расстояние AB достаточно велико, точка C начинает обгонять точку B [1], т. е. в сознание выводится наиболее вероятная траектория, хотя конкурирующие с ней версии еще не отброшены. Если в дальнейшем вероятность выводимой траектории становится ниже, чем у одной из конкурирующих, то возникает характерный сбой, который по психологическим проявлениям отождествляется с "юмористическим эффектом".

Таким образом, для наделения компьютера "чувством юмора" требуется решение "лингвистической проблемы" — распознавание последовательности многозначных образов — которая возникает также в исследованиях по автоматическому переводу [4, 5]².

² Несмотря на аналогию с проблемой машинного перевода, интересующая нас постановка задачи

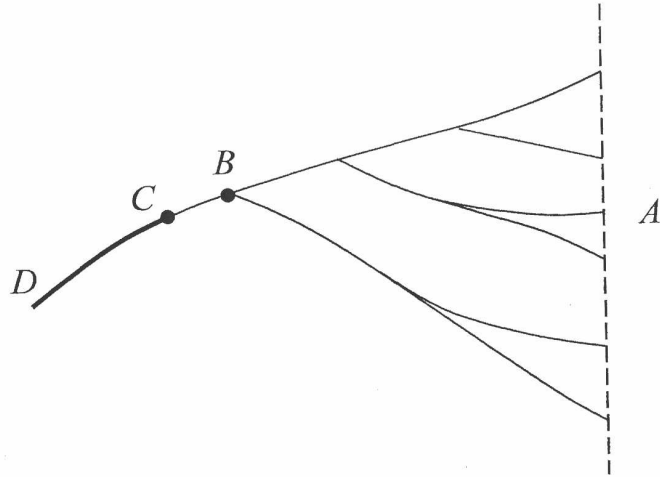


Рис. 2: Наглядное представление процесса обработки информации: тонкие линии — траектории, содержащиеся в оперативной памяти (или в подсознании), A — передний фронт, B — точка окончания ветвления, CD — участок траектории, выводимый на выход процессора (или в сознание).

Легко видеть, что она распадается на несколько более конкретных проблем:

- (1) составить список образов, имеющих хождение в популяции;
- (2) каждому слову изучаемого языка сопоставить набор ассоциированных с ним образов;
- (3) составить достаточно длинный текст для обучения, записанный непосредственно в образах;
- (4) сформулировать алгоритм обучения, сводящийся к корреляционному анализу сочетаемости образов;
- (5) на его основе выработать оценку вероятности для конечной последовательности образов.

Проблемы 1–3 в принципе тривиальны, но требуют громадного объема достаточно квалифицированной работы; трудно представить, чтобы кто-то занимался этим специально. Поэтому ниже мы рассмотрим возможность решения этих проблем в автоматическом режиме, ориентируясь на лингвистическую информацию, собранную в системах типа АВВУУ Lingvo. Проблемы 4, 5 более сложны: они не могут быть решены на умозрительном уровне и требуют длительного экспериментирования мето-

несколько отличается; эти отличия в основном и являются предметом дальнейшего обсуждения. Из трех стадий машинного перевода (анализ исходного текста, собственно перевод и генерация нового текста) для нас существенна лишь первая, но доведенная до максимально возможного уровня. В действующих программах эта стадия не является достаточно разработанной (см. Рис.б.1 в [4]).

дом "проб и ошибок". Мы предложим лишь предварительный вариант их решения, пытаясь подчеркнуть некоторые принципиальные моменты. Практический вариант алгоритма может быть сформулирован лишь с учетом имеющихся разработок для проблемы автоматического перевода [4, 5].

2. Идеальный язык как предельный случай

Предположим, что существует язык (пусть это будет латынь для определенности), который в некотором приближении может рассматриваться как идеальный. Под идеальным мы понимаем язык, для которого слова и образы находятся во взаимно-однозначном соответствии.

Тогда проблемы 1–3 решаются тривиально. Для того, чтобы составить список актуальных образов (проблема 1), достаточно переписать все латинские слова и пронумеровать их в алфавитном порядке. Если мы собираемся распознавать тексты на русском языке, то проблема 2 решается с помощью русско–латинского словаря: достаточно выписать все варианты перевода русского слова на латинский язык. В качестве текста для обучения (проблема 3) берется произвольный литературный текст на латинском языке. Обсудим теперь проблемы 4, 5.

Алгоритм обучения. Простейший способ обучения состоит в составлении корреляционной матрицы A_{ij} , где i и j пробегает номера образов. Перед началом обучения положим $A_{ij} \equiv 0$. Один шаг обучения будет состоять в анализе отдельного предложения, каждое из которых по определению выражает законченную мысль; поэтому все содержащиеся в нем образы (отождествляемые со словами в идеальном языке) являются взаимосвязанными. Для каждой пары слов (i, j) , входящих в данное предложение, увеличим на единицу элемент A_{ij} корреляционной матрицы (рис. 3),

$$\Delta A_{ij} = 1, \quad (1)$$

включая и случай $i = j$ (см. ниже). Разумеется, при таком способе обучения возможны ошибки: синтаксические связи в предложении имеют древесный характер (рис. 4), а существование ассоциативных связей гарантируется лишь между синтаксически связанными образами. Однако, совершенно некоррелированные образы попадают в одно предложение лишь с вероятностью $\sim (n/N)^2$ (n — число слов в предложении, N — число слов в языке), что нужно сравнить с вероятностью $\sim k(n/N)$ для ассоциативно-связанных образов (k — коэффициент корреляции). Ясно, что такая процедура позволяет эффективно обнаруживать корреляции между образами в широком диапазоне значений k ($1 \gtrsim k \gtrsim 10^{-3} \div 10^{-4}$).

Альтернативы такому способу обучения практически нет, так как синтаксический анализ можно проводить в автоматическом режиме лишь с большим процентом ошибок [4]. С другой стороны, основывать обучение только на синтаксических связях было бы тоже не очень правильно: предложения часто содержат описание повторяющихся

	1	2	3	4	5
1	1	1	0	1	0
2	1	1	0	1	0
3	0	0	0	0	0
4	1	1	0	1	0
5	0	0	0	0	0

Рис. 3: Изменение корреляционной матрицы при обработке предложения 4–2–1.

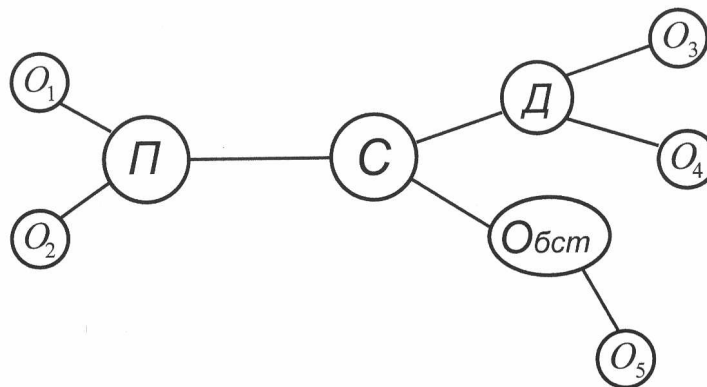


Рис. 4: Синтаксические связи в предложении имеют древесную структуру: подлежащее *П* связано со сказуемым *С*, которое в свою очередь связано с дополнением *Д* или обстоятельством *Обст*; к первому и последним могут иметься определения ($O_1 - O_5$).

ситуаций (напр. *Пастух гонит стадо*), в которых все образы являются ассоциативно связанными. В качестве последнего аргумента заметим, что при обучении человека синтаксический анализ, по-видимому, не играет большой роли: дети и малообразованные люди владеют разговорным языком вполне сносно, хотя и не имеют представления о синтаксисе.

Из приведенных оценок ясно, что обучение желательно проводить на коротких предложениях — с уменьшением n ошибка обучения уменьшается. Повышается также и общая эффективность процесса: при обработке двух предложений из 10 слов требуется рассматривать $10^2 \cdot 2 = 200$ попарных связей, тогда как при анализе одного предложения из 20 слов таких связей $20^2 = 400$. Отсечение же длинных предложений не требует существенной затраты времени и ресурсов.

Вероятность траектории. После того, как корреляционная матрица A_{ij} сформирована в результате обучения на достаточно длинном тексте, вероятность некоторой конечной последовательности образов можно определить как ³

$$p = \sum_{i,j (i \neq j)} A_{ij}, \quad (2)$$

где i и j пробегают номера образов, задействованных в данной последовательности. В основу определения вероятности положен чисто аналоговый принцип, так что наиболее вероятными оказываются комбинации образов, которые часто встречаются в обучающем тексте. В определении (2) мы исключили члены с $i = j$, так как интересующая нас вероятность должна характеризовать связность траектории.

На первый взгляд, аналоговый характер алгоритма требует, чтобы условие $i \neq j$ было введено и в правило обучения (1). В действительности это не так, поскольку обучение и распознавание производятся в разных условиях: для обучения мы используем замкнутые предложения, тогда как распознаванию подвергается произвольный отрезок текста, независимо от границ предложения. Если в (1) наложить условие $i \neq j$, то самокорреляция образов окажется почти нулевой: повторное употребление слова в коротком предложении считается стилистической ошибкой ⁴ и практически не встречается в литературном тексте. Из общих же принципов ясно, что корреляция образа самого с собой является максимальной, что и обеспечивается включением случая $i = j$ в правило обучения (1). В процессе распознавания такая самокорреляция образов имеет существенное значение: связный текст обычно характеризуется наличием "главного героя", образ которого присутствует почти в каждом предложении. Поэтому наличие одинаковых образов в соседних предложениях является вполне типичным и их самокорреляция существенна для адекватной оценки связности текста.

³ Нормировка вероятности оказывается произвольной, но это несущественно для сравнения траекторий.

⁴ Если это не артикль или служебное слово.

3. Реальный язык: никто не хотел "как хуже".

Разумеется, любой реальный язык очень далек от идеального: с одной стороны, каждое слово ассоциируется со множеством образов, с другой стороны, один образ может выражаться разными словами. Возникает впечатление, что вмешался какой-то злой дух и специально испортил все существующие языки⁵. На самом деле, никто не хотел "как хуже" и никто не ставил целью специально запутать ситуацию: неоднозначность актуальных языков является естественным следствием их развития.

Первые слова древнего человека формировались по принципу "что вижу, то пою"⁶: напр. слова *лиса*, *волк*, *медведь* возникли из восклицаний охотников, предупреждающих соплеменников о появлении соответствующего животного. Естественно, эти слова четко ассоциировались с указанными образами и не сопровождались никакой неоднозначностью (рис. 5,а). Усложнение взаимоотношений между людьми сделало актуальными вопросы социального поведения и привело к появлению слов типа *характер*, *манеры*. Комбинирование этих слов с уже существующими привело к возникновению составных понятий: *лисий характер*, *волчий характер*, *медвежьи манеры*; эти понятия оказались плодотворными и для их обозначения появились специальные слова — *хитрость*, *жестокость*, *неуклюжесть* и т. д. (рис. 5,б). Нетрудно видеть, что язык адекватно отреагировал на изменение ситуации: возникли новые образы — появились новые слова; при этом количество слов осталось равным количеству образов, актуальных для популяции (рис. 5,б). Однако, уже на этом этапе возникает запутанность языка: с одной стороны, появилась синонимическая неоднозначность (можно сказать *хитрость*, а можно — *лисьи манеры*), с другой — старые слова приобрели новые значения (слово *лиса* стало обозначать не только "рыжую с хвостом", но и хитрую Марью Ивановну). Мы видим, что запутанность языка является неизбежным следствием его развития: новые образы сначала описываются существующими словами, а затем для них придумывают специальные названия; но ассоциативных связей со старыми словами никто не отменяет.

К счастью, запутанность языка, возникающая в результате его развития, является легко устранимой. У каждого слова можно выделить его главное значение (сплошные стрелки на рис. 5,б и ниже) и второстепенные значения (пунктирные стрелки). Если сопоставлять каждому слову образ, ассоциированный с его главным значением, то взаимно-однозначное соответствие между словами и образами восстанавливается (рис. 5,б).

К сожалению, неоднозначность языка возникает и по другим причинам, которые с точки зрения самого языка являются внешними. Так, если в двух провинциях одно и то же понятие выражалось разными словами, то при объединении этих провинций в одно государство приходится признавать допустимыми оба слова. Это приводит к возникновению неприводимых синонимов (рис. 6,а), под которыми мы понимаем слова с одинаковым главным значением, в отличие от приводимых (рис. 6,б), для которых

⁵ Легенда о Вавилонской башне возникла, по-видимому, под влиянием этого впечатления.

⁶ Обычай некоторых азиатских племен.

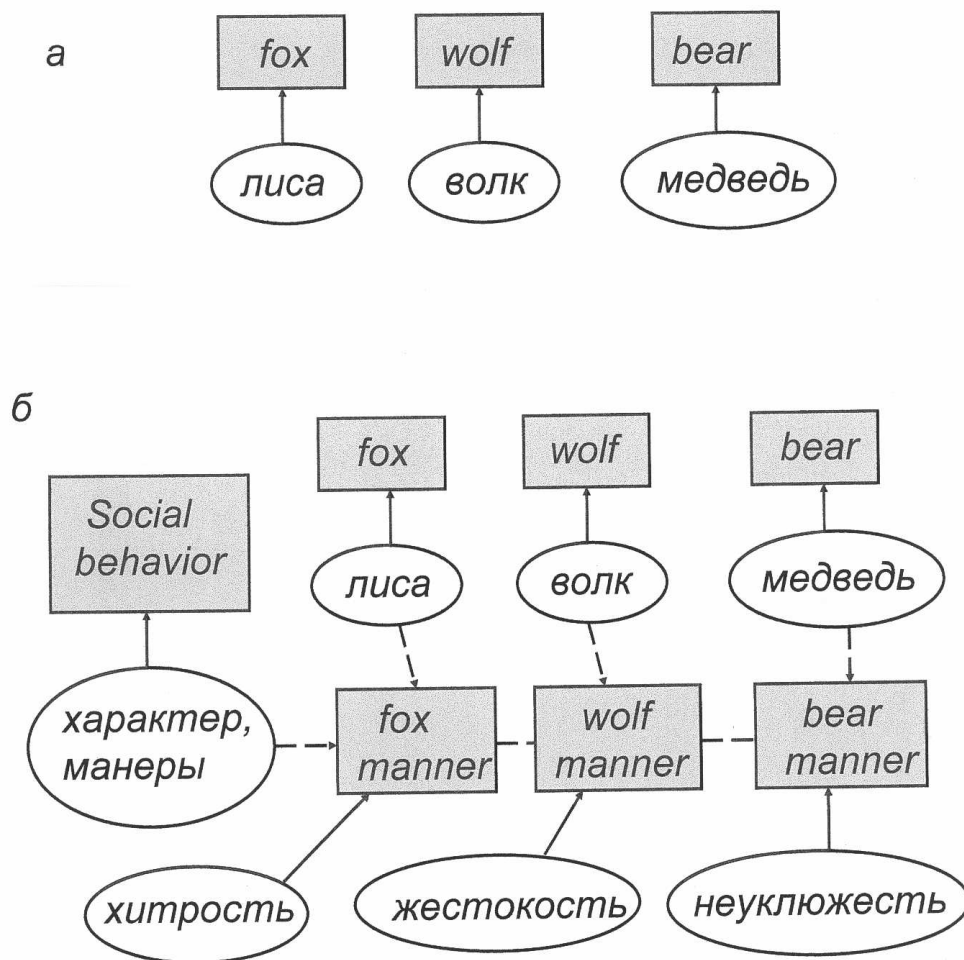


Рис. 5: а — Первый этап развития языка: слова и образы находятся во взаимно-однозначном соответствии. б — Следующий этап развития языка: показан основной механизм возникновения его неоднозначности. Для большей ясности образы обозначены английскими словами.

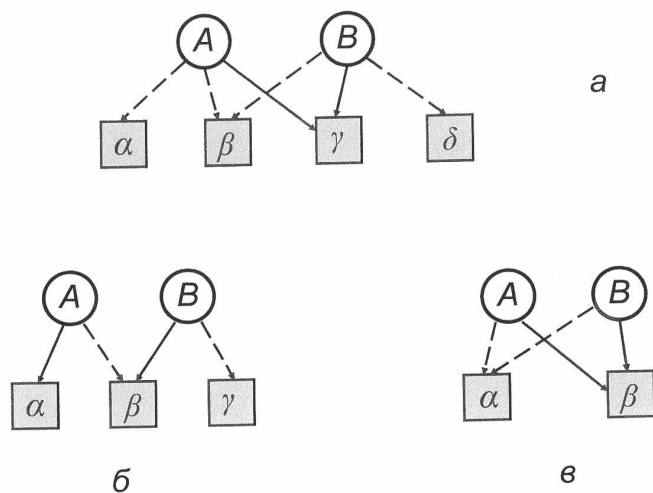


Рис. 6: (а) Слова A и B — неприводимые синонимы, их главные значения соответствуют одному образу γ . (б) Слова A и B являются синонимами по отношению к образу β , но они приводимы, так как их главные значения различны. (в) Для полноценных синонимов совпадают как главные, так и второстепенные значения.

главные значения различны. К тому же эффекту приводит социальное расслоение общества: так сексуально окрашенные образы для малообразованной части населения ассоциировались с нецензурными словами; с другой стороны, для них существуют нейтральные наименования (возникшие в аристократической среде) и медицинские термины латинского происхождения. Тот же механизм может действовать и в обратную сторону: если одни и те же слова использовались для обозначения принципиально различных объектов (напр. слово *коса* в русском языке), то возникают омонимы, т. е. слова с несколькими главными значениями.

Неприводимые синонимы и омонимы фактически являются дефектами языка; они возникли по случайным причинам и их существование не оправдано никакими разумными соображениями.

4. Реальный язык в качестве идеального.

Проведенный в разд. 3 анализ позволяет понять, какие коррекции нужно внести в алгоритм, если в качестве идеального (латынь в разд. 2) мы попытаемся использовать реальный язык (напр. немецкий).

Чтобы получить список образов, достаточно (в первом приближении) переписать все немецкие слова и ассоциировать их с главными значениями (т. е. слово считается

символическим обозначением образа, который соответствует его главному значению). В первом приближении сохраняется и весь алгоритм разд. 2: набор образов, ассоциируемых с русским словом, получается с помощью русско-немецкого словаря, для обучения используются немецкие тексты и т. д. Главная ошибка, которая при этом совершается, состоит в том, что корреляционные связи реальных образов подменяются корреляцией немецких слов. Однако, с этой ошибкой по-видимому нужно смириться, так как человек поступает точно так же. В человеческой практике не принято составлять списки образов, тогда как списки слов (словари) легко доступны. Не существует и достаточно длинных текстов, записанных в образах, но доступны длинные тексты, записанные в словах (книги). Поэтому при обучении человека неизбежно происходит подмена образов на обозначающие их слова. Сравнительно небольшая погрешность такого обучения связана со следующим:

(а) Для большинства слов их "главные" значения являются действительно главными, т. е. слово с подавляющей вероятностью употребляется в этом значении.

(б) Второстепенные значения обычно логически связаны с главными (см. рис. 5,б), поэтому корреляции всего конгломерата значений с другими образами не сильно отличаются от корреляций для главного значения.

(в) В человеческой практике принято проводить обучение на некотором стандартном наборе "классических" текстов. При этом подразумевается, что "классики" выражают свои мысли более четко, чем большинство населения, т. е. употребляемый образ как правило обозначается словом, имеющим его в качестве главного значения. Так или иначе, на одинаковых текстах люди обучаются примерно одинаково (даже если и не совсем правильно), что способствует их взаимопониманию.

Теперь посмотрим, какие поправки на неидеальность языка можно реально сделать.

Неприводимые синонимы. Наличие неприводимых синонимов приводит к тому, что в составленном нами списке образов некоторые образы будут повторяться по несколько раз. Для упорядочения ситуации удобно исходить из концепции, что полноценных синонимов (рис. 6,в) практически не бывает: даже если в какой-то момент слова A и B были эквивалентны (для них совпадали как главные, так и второстепенные значения), то с течением времени они расходятся, по-разному обрастая новыми значениями (рис. 6,а), и начинают употребляться в различных контекстах⁷. С этой точки зрения, неприводимые синонимы отмечают некоторые оттенки основного значения, т. е. соответствуют близким, но не совпадающим образам.

Сочетаемость этих слегка различных образов с другими образами адекватно описывается корреляционной матрицей, получаемой по правилу обучения (1) (поскольку фактически корреляции изучаются между словами). Проблемы связаны лишь с корреляцией между самими синонимами и аналогичны тем, что возникают в связи с самокорреляцией образов (разд. 2). Одновременное употребление двух синонимов в

⁷ Для ситуации на рис. 6,а: если в некотором контексте допустимо появление образов α и γ , то образ γ должен обозначаться словом B ; употребление слова A приводит к реальным недоразумениям.

коротком предложении столь же нежелательно, как и повторное употребление одного и того же слова. Поэтому описанная в разд. 2 процедура обучения приведет к практическому отсутствию корреляций между близкими образами, тогда как фактически эти корреляции очень сильны.

В качестве модели неприводимых синонимов можно принять, что появляющийся образ с вероятностью p_A обозначается словом A , с вероятностью p_B — словом B , с вероятностью p_C — словом C и т. д.⁸ Тогда легко показать (см. приложение), что блок корреляционной матрицы, соответствующий синонимам $A, B, C \dots$ должен иметь структуру

$$\begin{pmatrix} S_{AA} & \sqrt{S_{AA}S_{BB}} & \sqrt{S_{AA}S_{CC}} & \dots \\ \sqrt{S_{BB}S_{AA}} & S_{BB} & \sqrt{S_{BB}S_{CC}} & \dots \\ \sqrt{S_{CC}S_{AA}} & \sqrt{S_{CC}S_{BB}} & S_{CC} & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}. \quad (3)$$

При обучении согласно алгоритму (1) диагональные элементы $S_{AA}, S_{BB}, S_{CC} \dots$ определяются правильно, тогда как недиагональные оказываются практически нулевыми; но их можно поправить искусственно, в соответствии с матрицей (3). Таким образом, "обучение синонимам" производится специально, как это имеет место и в человеческой практике.

Омонимы. С лингвистической точки зрения, омонимы считаются разными словами, что прямо отмечается в словарях (как правило, римскими цифрами, напр. *коса I, коса II* и т. д.). Поэтому при составлении списка образов с омонимами не связано никаких проблем и они фиксируются как разные образы. Однако, на письме и в речи омонимы не различаются, что создает проблемы при обучении.

Применительно к рассматриваемой ситуации это означает, что "обучение омонимам" должно производиться вручную: встретив в обучающем тексте один из зарегистрированных омонимов, компьютер должен запрашивать оператора, который из них подразумевается в данном предложении. Однако, такой ручной этап может быть не очень продолжительным: после того, как для корреляционной матрицы получена минимальная статистика, идентификацию омонимов можно доверить компьютеру: они, как правило, употребляются в существенно различном контексте и их ассоциативные связи резко различаются.

Отсутствие образов. В составленном нами списке образов некоторые образы могут просто отсутствовать, если для их обозначения не придумано специальных слов (как правило, это относится к новым или малоупотребительным образам). Последнее означает, что образ еще не вполне прижился в популяции, и многим людям он попросту неизвестен; поэтому незнание его можно простить и компьютеру.

Разумеется, такие образы можно внести в список вручную, сопоставляя его с некоторой группой слов. Но тогда вручную придется проводить и обучение: компьютер

⁸ Такая модель исходит из того, что различия между близкими образами носят символический характер, и в человеческой практике на них не обращают внимания.

должен запрашивать, соответствует ли найденная им группа слов указанному образцу или эти слова попали в одно предложение случайно.

5. Заключение.

Выше предложен предварительный вариант решения проблем 1–5, сформулированных во введении. При рассмотрении первых трех проблем мы ориентировались на лингвистическую информацию, содержащуюся в системах типа АBBYU Lingvo. В таких системах четко различаются главное значение слова (оно фиксируется как "первое значение") и его второстепенные значения, а также дается список его синонимов. Последние, впрочем, нужно проверять на приводимость, но алгоритм такой проверки ясен из вышеизложенного. Система АBBYU Lingvo содержит также большое число фразеологических сочетаний, многие из которых соответствуют оригинальным образам; к сожалению, их выделение требует дополнительной работы.

Предложен также аналоговый алгоритм обучения и распознавания (проблемы 4, 5), который должен рассматриваться как предварительный и доводиться до реалистичного уровня экспериментальным путем, на основе существующих программ для машинного перевода [4, 5]. Можно надеяться, что проведенный анализ дает достаточный материал для начала таких экспериментов, которые могут сделать реализацию компьютерного "чувства юмора" делом ближайшего будущего.

Приложение. Корреляция неприводимых синонимов.

Как ясно из текста, для неприводимых синонимов используется модель, согласно которой появляющийся образ с вероятностью p_A обозначается словом A , с вероятностью p_B — словом B , с вероятностью p_C — словом C и т. д. Повторное появление одного образа в коротком предложении является маловероятным, и типичная ситуация соответствует наличию двух одинаковых образов в соседних предложениях. Вероятности конфигураций AA , AB , ... равны соответственно p_A^2 , $p_A p_B$, ..., что приводит к корреляционной матрице

$$\begin{pmatrix} p_A^2 & p_A p_B & p_A p_C & \dots \\ p_B p_A & p_B^2 & p_B p_C & \dots \\ p_C p_A & p_C p_B & p_C^2 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}, \quad (4)$$

которая отличается от матрицы (3) постоянным множителем, ввиду произвольности нормировки последней. Диагональные элементы S_{AA} , S_{BB} , S_{CC} ... матрицы (3) можно считать известными, так как они правильно определяются при обучении по алгоритму разд. 2; недиагональные же элементы определяются из соответствия (3) с (4).

Список литературы

- [1] И.М.Суслов, Биофизика **37**, 318 (1992).⁹
- [2] И.М.Суслов, Биофизика **37**, 325 (1992).
- [3] И.М.Суслов, Компьютерная хроника, 1994, вып.1
- [4] Hutchins W.J., Harold L.S. An Introduction to Machine Translation, London, Acad.Press, 1992.
- [5] Исчерпывающую библиографию по проблеме машинного перевода можно найти на сайте Дж.Хатчинса <http://www.hutchinsweb.me.uk>

⁹ Работы [1, 2, 3] можно найти на авторском сайте (<http://www.kapitza.ras.ru/~suslov>).