

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ (Вычислительные системы)

2002 год

Выпуск 171

УДК 519.769 : 801.314.4

ЯЗЫК ЗАГОЛОВКОВ: КОЛИЧЕСТВЕННЫЙ АНАЛИЗ ПРОТОТИПОВ И СХЕМ ВАРЬИРОВАНИЯ ¹

В.Д. Гусев, Н.В. Саломатина

В в е д е н и е

Объектом исследования в данной работе являются заголовки, построенные по принципу варьирования хорошо известных (находящихся “на слуху”) прототипов, в качестве которых могут выступать строки популярных песен, названия литературных произведений, картин, кинофильмов, устойчивые словосочетания (поговорки и пословицы, фразеологизмы, термины, лозунги), высказывания (и даже фамилии) известных людей (политиков, общественных деятелей, писателей, актеров) и т.д. Заголовки подобного типа, часто основанные на “игре слов”, составляют заметную долю от общего их количества, причем наблюдается устойчивая тенденция к ее повышению.

Схемы варьирования, используемые при переходе от прототипа к заголовку, весьма разнообразны, но вполне обозримы. Многие из них на качественном уровне описаны в [1, 2]. Некоторые схемы варьирования могут быть формализованы уже на данном этапе, что позволяет в перспективе создать автоматизированную систему, моделирующую с той или иной степенью достоверности деятельность автора при выборе заголовка. Она должна обеспечивать поиск возможных прототипов для заголовков и их (пусть и ограниченное) варьирование.

¹Работа выполнена при финансовой поддержке РФФИ, проект 00-06-80420.

Для выработки приоритетов при создании такой системы важно иметь не только качественные, но и количественные характеристики состава потенциально возможных прототипов и способов перехода от прототипа к заголовку. В соответствии с этой целью работы является классификация прототипов, используемых при подборе заголовка, и схем их варьирования, а также ранжирование тех и других по частоте встречаемости. Материалом для исследования послужила собранная авторами за период с 2000-го по 2002-й год подборка из нескольких тысяч газетных заголовков, построенных по принципу варьирования известного прототипа. Начальный этап работ по данной тематике отражен в [3].

1. Формирование подборки заголовков

Степень использования скрытого (варьированного) цитирования в заголовках газет сильно отличается у разных изданий. Наша подборка, в основном, получена на материале газет "Известия" (И), "Труд" (Т), "Аргументы и факты" (АиФ), "Повск" (П) (2000 ÷ 2002 г.г.), в которых применение цитирования, особенно построенного на "игре слов", носит достаточно систематический характер.

Для каждого заголовка указывается: 1) соответствующий ему прототип; 2) источник, откуда взят прототип; 3) подзаголовок (если он присутствует); 4) координаты издания (название газеты, год, номер выпуска, номер страницы); 5) фамилия, имя, отчество автора статьи. Отметим, что прототипов может быть несколько (обычно не более двух) в случае составных заголовков типа "Продажные шкуры неубитых медведей" (И, 2000 г., 53, с. 5). Здесь фигурируют два прототипа: "Продажная шкура" и "Шкура неубитого медведя", что подтверждается подзаголовком: "Выгодная торговля Родяной вообще не преступление государства, а его доблесть...".

Подзаголовок также может быть два. Второй подзаголовок вводился нами в ситуациях, когда первый подзаголовок (авторский) не объяснял выбор прототипа для названия статьи и характер его варьирования. Такие случаи не так уж редки, что

отмечал еще В.Г.Костомаров в [4]. К сожалению, далеко не всегда заголовок, содержащий скрытое цитирование, сопровождается авторским подзаголовком. В этом случае функции автора должны взять на себя составители подборки, иначе элементы своеобразной словесной игры, демонстрируемые автором статьи, могут оказаться незамеченными (или непонятыми) читателем.

Объективные трудности возникают и при идентификации источника, откуда взят прототип. В сомнительных случаях мы обращались к электронной энциклопедии и к имеющимся в нашем распоряжении справочным изданиям, но это не всегда приводило к успеху, т.е. некоторая (незначительная) часть прототипов осталась неидентифицированной. Возможны также случаи неоднозначной идентификации, когда какой-нибудь прототип, например, "Вся королевская рать", может быть идентифицирован как название романа Роберта П. Уоррена или как название представленного по нему кинофильма.

В подборку включались не только заголовки, имеющие прототип, но и не имеющие его, однако, структурно похожие на известные прототипы. Это сходство может выражаться в наличии *повторности* ("Дядя Паша Бородин не останется один" (И, 23.02.2001, с.2), "Стихия становится стихийнее" (И, 21.03.2001, с.2), "И тут и там звучал тамтам" (II, 19, 1999.)), *использовании паронимов* ("Скетч о скотче" (И, 23.02.2001, с.6)), *антонимов* ("При наличии отсутствия" (И, 7.07.2001, с.1)) и других выразительных средств и приемов. Дополнительным аргументом в пользу включения подобных заголовков в подборку являлось наличие при каждом из них поясняющего подзаголовка. Этот факт можно рассматривать как авторское указание на присутствие элементов "языковой игры", требующей расшифровки. Например, заголовок "Бескомпромиссный компромисс" (И, 6.06.2001, с.4, Е. Григорьева) сопровождается подзаголовком "Военнослужащим будут платить больше, но позже", четко и лаконично обосновывающим выбор конструкции заголовка.

Из сказанного ясно, что формирование подборки заголовков, построенных на игре слов, не является чисто технической проблемой. Подборка, сформированная на основе предложенной нами структуры данных, может в дальнейшем служить не

только в качестве источника возможных прототипов и образцов варьирования, но и для исследовательских целей (ранжирование прототипов по частоте использования, выявление индивидуальных (авторских) особенностей варьирования, поиск неявных заимствований, дифференциация печатных изданий по используемым прототипам и схемам варьирования и т.п.).

2. Качественная и количественная характеристика прототипов и их источников

Создаваемая нами база данных, призванная обеспечить компьютерную поддержку процедуры поиска нужного прототипа для заголовка, будет тем эффективнее, чем больше ее объем. Пополнять эту базу новыми прототипами можно, не только анализируя газетные заголовки, но и пользуясь другими источниками (различными словарями, сборниками популярных песен, литературными произведениями и т.п.). Чтобы эта деятельность была более целенаправленной, желательно на основе уже имеющейся подборки получить количественные оценки использования прототипов разного вида.

В табл.1 приведены количественные характеристики встречаемости прототипов разного вида в анализируемой подборке. Выделены шесть основных групп источников, из которых черпают прототипы для заголовков авторы публикаций (см. столбцы 1 и 2), и седьмая дополнительная со значением "все остальное", разнородная по своему составу. Группы расположены в порядке убывания своей "значимости" (в смысле частоты обращения к ним как к источникам возможных прототипов).

В третьем столбце проводится дифференциация прототипов внутри каждой из выделенных групп (за исключением групп 6 и 7). Она не вызывает особых затруднений для групп 2 ÷ 5; некоторые неоднозначности могут возникнуть лишь внутри группы 1. Они обусловлены отсутствием у лингвистов единого мнения о том, что такое фразеологизм, что выделяет его из множества устойчивых сочетаний слов, следует ли относить к фразеологизмам пословицы и поговорки и т.д. (обсуждению этих вопросов уделено значительное место во вводной части к известному фразеологическому словарю А.И. Молоткова [5]). Не вдаваясь в дискуссию по этому вопросу, отметим лишь, что мы при

Абсолютные (F) и относительные ($F(\%)$) частоты встречаемости прототипов
разного вида в анализируемой подборке

№	Источник	Характеристика	F	$F(\%)$	Примеры
1	Тематические подборки, словари, справочники, учебники и т.п. $F_{\Sigma} = 904$ (42,9%)	пословицы, поговорки	439	20,8%	Клин клином вышибают Лес рубят — щепки летят Язык мой — враг мой
		фразеологизмы	166	7,9%	Без царя в голове Вопрос ребром Дело в шляпе
		термины, факты, утверждения	96	4,5%	Комплексы неполноценности Волга впадает в Каспийское море От перемены мест слагаемых сумма не меняется
		устойчивые словосочетания	78	3,7%	Брак по расчету Закрома Родины Гадание на кофейной гуще
		объявления, лозунги, призывы, реклама (слоганы)	61	2,9%	Пьянству — бой! Чистота — залог здоровья Голосуй или проиграешь
		цитаты из библии	50	2,4%	Не судите, да не судимы будете Не хлебом единым Возвращение блудного сына
		жаргонизмы	14	0,7%	Крыша едет Кина не будет Хорошо сидим
2	Литературно- публицистические произведения $F_{\Sigma} = 416$ (19,7%)	названия художественных произведений, литературные цитаты	191	9,0%	Бегущая по волнам На деревню дедушке К нам едет ревизор
		названия стихотворных произведений, стихотворные строки	185	8,8%	Я достаю из широких штанин... Отсель грозить мы будем шведу Кому на Руси жить хорошо
		сказочные присловья, правоучения	28	1,3%	Не пей из копытца — козленочком станешь По пучьему веленью
		цитаты из публицистических произведений	12	0,6%	Единство и борьба противоположностей Учиться, учиться и еще раз учиться

	Источник	Характеристика	F	F(%)	Примеры
3	Музыкальные произведения $F_{\Sigma} = 338$ (16,0%)	названия и строки из: песен	306	14,5%	Не нужен нам берег турецкий Броня крепка и танки наши быстры Вагончик тронется, перрон останется
		романсов	14	0,7%	Отцвели уж давно хризантемы в саду И звезда с звездой говорит
		других произведений	18	0,8%	Ария варяжского гостя Сказки венского леса
4	Кино, радио, ТВ $F_{\Sigma} = 170$ (8,1%)	названия и "ходячие выражения" из: художественных фильмов	144	6,8%	Особенности национальной охоты О бедном гусаре замолвите слово
		ТВ- и радиопередач	14	0,7%	Говорит и показывает Москва
		мультфильмов	8	0,4%	Ребята, давайте жить дружно
		других источников	4	0,2%	
5	Авторские высказывания, цитаты дня, афоризмы $F_{\Sigma} = 142$ (6,7%)	известные изречения: политиков	45	2,1%	Хотели как лучше, получилось как всегда
		писателей	40	1,9%	Создать мир легче, чем понять его
		ученых	17	0,8%	Российское могущество прирастает будет Сибирью
		философов	8	0,4%	Все течет, все изменяется
		других лиц	24	1,1%	Велика Россия, а отступать некуда
		неидентифицированные высказывания	8	0,4%	
6	Газеты, журналы $F_{\Sigma} = 114$ (5,4%)	заголовки, не имеющие прототипов, но структурно сходные с ними	14	5,4%	Кино, вино и домино "Мина" Минфина Новый старый завод
7	Другие источники $F_{\Sigma} = 25$ (1,2%)	названия картин, скульптурных произведений, считалки, загадки и т.п.	25	1,2%	Переход Суворова через Альпы Купание красного коня Ехал Грека через реку "А" и "Б" сидели на трубе

ПРИМЕЧАНИЕ. F_{Σ} — суммарная абсолютная частота встречаемости в подборке прототипов из данного класса источников.

определении фразеологизмов старались придерживаться классификации акад. В.В. Виноградова (фразеологические сращения и единства), т.е. понимали их как идиоматические обороты.

Пятый столбец ($F(\%)$) дает возможность упорядочить все категории прототипов по относительной частоте представленности их в нашей подборке. На первом месте стоят пословицы и поговорки (20,8%), на втором — названия и стихотворные строки из песен (14,5%), далее идут названия художественных произведений, цитаты из них, образные выражения (9%), потом названия стихотворений и отдельные стихотворные строки (8,8%), фразеологизмы (7,9%) и, наконец, названия известных кинофильмов и “ходячие выражения” из них (6,8%). Полученная статистика определяет наиболее перспективные источники для пополнения подборки прототипов и порядок обращения к ним.

Заголовки, не имеющие прототипов, но структурно сходные с ними (см. группу 6 и примеры в разделе 1) также составляют заметную долю от общего объема подборки (5,4%). От “истинных” прототипов они отличаются тем, что пока еще не находятся “на слуху”. Однако, нередко в них просматриваются новые тенденции в плане построения заголовка и пояснения его.

В заключение данного раздела заметим, что интуитивные ощущения о частоте использования того или иного класса прототипов или приемов их варьирования не всегда согласуются с реальными цифрами. Так, Е.А. Земская в [1] на первое место по частоте использования в качестве прототипа ставит стихотворные строки, упорядочивая авторов следующим образом: Пушкин, далее Лермонтов, Грибоедов, Маяковский. В нашей статистике стихотворные строки стоят лишь на четвертом месте и порядок авторов иной. Маловероятно, что десятилетний интервал, разделяющий две работы, мог явиться причиной столь существенных различий.

3. Качественная и количественная характеристика схем варьирования прототипов

3.1. *Классификация схем варьирования.* Существуют различные и лишь частично согласующиеся друг с другом классификации схем варьирования устойчивых словосочетаний вообще

и прототипов для газетных заголовков в частности. При построении нашей классификации мы в существенной степени опирались на результаты работ [1, 2, 6]. Две первые посвящены непосредственно газетным заголовкам; последняя же, по утверждению авторов, представляет единственный в мировой практике (по состоянию на 1997 г.) “опыт описания идиом и пословиц в их вариантном многообразии и речевой динамике”. В связи с этим авторы [6] уделяют значительное внимание определению основных типов трансформаций фразеологических единиц в тексте.

Отметим две особенности нашей классификации:

1) она носит “операциональный” характер в том смысле, что различие между заголовком и его прототипом определяется как минимальное число операций заданного типа, переводящих прототип в заголовок. Базовыми при этом являются операции замены, вставки и устранения отдельных слов или групп слов (в последнем случае операции называются блочными или множественными). В основе такого подхода лежит понятие “редакционного расстояния” [7]. С его помощью удобно алгоритмизировать процедуру выявления ближайших соседей заданной языковой единицы. Данный подход использовался нами для анализа вариативности корней слов и канонических форм слов русского языка (см. [8, 8]). Семантические аспекты учитываются, по возможности, при детализации типов замен, вставок и т.п.

2) классификация проводилась параллельно с оцениванием частоты использования того или иного типа варьирования. Это позволяет устранять из рассмотрения редко встречающиеся схемы преобразования прототипа, т.е. не перегружать классификацию излишними деталями. С другой стороны, излишне частое посещение какого-либо узла классификационной схемы в процессе набора статистики может служить основанием для его расщепления на ряд подузлов.

В приводимой ниже классификации звездочками отмечены пункты, не встречающиеся в явном виде в классификациях из [1, 2, 6]. Частоты встречаемости на практике различных схем варьирования прототипов оцениваются лишь в данной работе. Для удобства они сведены далее в табл. 2. В иллюстрирующих примерах заголовков выделяется кавычками, прототип — знака-

ми < ... >; подзаголовок — двумя косыми чертами (приводится не всегда).

Классификация имеет следующий вид:

- 0 — изменение грамматических характеристик одного из компонентов прототипа.

0.1* — отсутствие варьирования (заголовок повторяет прототип): “Обещанного три года ждут” (И, 2000, № 157); / Банк России выпустит тысячерублевую банкноту/;

0.2 — изменение числа, рода: “Незнайки на Луне” (Т, 2001, № 8); <Незнайка на Луне> — повесть Носова; /Вместо науки Россия, похоже, станет ублажать в космосе богатых ротозеев/; “Принц и нищий” (И, 19.06.2001); <Принц и нищий> — повесть М.Твена; /Описание двух биографий — преуспевающего бизнесмена и медсестры горбольницы/;

0.3 — изменение видо-временных значений глаголов: “Никто не хочет умирать” (Т, 2002, № 11); <Никто не хотел умирать> — название кинофильма; /Наступает момент истины для российских нефтяных компаний/;

0.4 — изменение значений лица местоимения и/или глагола: “Широка страна его родная” (И, 2000, № 128); <Широка страна моя родная> — строка из песни; /Газпром больше не будет московской компанией/;

- 1 — замена одного слова другим, обычно связанным с содержанием статьи. Заменяющее слово часто фигурирует в подзаголовке. Будем различать замены с использованием:

1.1 — антонимов: “Дорого и сердито” (И, 2000, № 245); <Дешево и сердито>; /Сколько будет стоить автомобиль/;

1.2 — *слово из того же таксономического класса* (включая замену одного собственного имени другим): “Сирота украинская” (И, 2000, № 115); <Сирота казанская>; /Компания Нефтегаз Украины лишилась своего временного начальника/; “Мистер Барнет в тылу врага” (АиФ, 2002, № 5) <Мистер Питким в тылу врага>; /В Новосибирске состоялся премьерный показ фильма “В тылу врага”/;

1.3 — *паронимов в широком смысле* (близкие по звучанию слова, омофоны, омоформы): “Тени забытых предков” (И, 2000, № 97); <Тени забытых предков>; /Коллегия Верховного суда отказалась снять с Берин расстрельную статью/; “Кассовый фильм” (И, 2000, № 246) <Кассовый фильм>; /Режиссер Клод Лелюш доверил певице Патрисии Каас роль в своей новой картине/; “ЕДИНСТВО и борьба противоположностей” (И, 2000, № 184); <Единство и борьба противоположностей>; /Выдержат ли медведи испытание региональными выборами .../;

1.4 — *многозначный слог и омонимов* (в данном случае речь идет об изменении значения слова, но не формы): “Губернатор Петербурга остепенился” (И, 25.04. 2001); /Вчера губернатор Петербурга Владимир Яковлев стал доктором экономических наук/; “Нам любимые до’роги доро’ги” (АиФ, 2001, № 41); <Нам любимые до’роги доро’ги> — строка из песни; /На что тратятся средства дорожного фонда/;

1.5 — *слово из того же лексико-грамматического класса*: “Зачем вы, девушки, реонимыз любите” (Т, 2002, № 12); <Зачем вы, девушки, красивые любите>; /30-летний житель Сальска подарил мягкую игрушку, начиненную взрывчаткой, своей подруге.../; “Рейтинги по осени считают” (АиФ, 2001, № 39); <Цыплят по осени считают>; /Осень — это как начало нового года, только политического/;

• 2 — *вставка слова:*

2.1 — *прилагательного*: “Сеанс одновременной политической игры” (И, 2000, № 154); <Сеанс одновременной игры>; /Житель Казани выдвигается в президенты сразу двух соседних республик/;

2.2 — *существительного*: “Удар каской ниже пояса” (И, 2000, № 95); <Удар ниже пояса>; /В мае традиционно расцветают все цветы профсоюзного движения/;

2.3 — *других частей речи;*

• 3 — *устранение слова:*

3.1 — *существительного*: “Летайте самолетами” (И, 9.02.2001); <Летайте самолетами Аэрофлота>; /Сегодня, 9 февраля — день рождения Аэрофлота.../;

3.2 — *глагола*: “На чистую воду” (И, 2000, № 202); <Вывести на чистую воду>; /За каждого браконьерски добытого рака нарушителей ждут штрафные санкции/;

3.3 — *других частей речи;*

• 4 — *изменение синтаксической структуры прототипа, его направленности, роли отдельных компонентов:*

4.1. — *перевод утвердительных форм в отрицательные и наоборот*: “Третий не лишний” (И, 6.06.2001); <Третий лишний>; /Кремль оставит губернаторам право на дополнительный (третий) срок/; “В ногах правда есть” (Т, 2001, № 12); <В ногах правды нет>; /В городке Рассказово Тамбовской области все вяжут шерстяные носки/;

4.2 — *смена интонации* (обычно выражается заменой точки на вопросительный знак и наоборот): “Свой среди чужих, чужой среди своих?” (Труд-7,

2001, № 40); <Свой среди чужих, чужой среди своих.> — название фильма; /Первому и последнему Президенту СССР исполняется 70 лет/;

4.3* — *частичное отрицание*: “Бывает дым и без огня”; (И, 2000, № 171); <Нет дыма без огня>; /Китайские медикаменты напугали японцев/;

4.4 — *изменение синтаксической структуры и роли компонентов* (часто реализуется заменами с использованием слов из других лексико-грамматических классов): “Гори, мотор, ясно, чтобы не погасло” (И, 2001, № 121); <Гори, гори ясно, чтобы не погасло>; “Преступление без наказания” (И, 2000, 144); <Преступление и наказание>; /Саддам Хусейн до сих пор считает, что нападение на Кувейт было оправдано/;

• 5* — *двойная замена*:

5.1* — *рядом стоящих слов с сохранением синтаксической структуры прототипа*: “Хорошее кресло пусто не бывает” (И, 2000, № 201); <Свято место пусто не бывает>; /Рынок мягкой мебели захватили кустари/; “Место преступления определить нельзя” (Труд-7, 2001, № 35); <Место встречи изменить нельзя>; /Только за последние два года число правонарушений в сфере высоких технологий выросло в России в 20 раз/;

5.2* — *разнесенных слов с сохранением синтаксической структуры прототипа* (при этом разнесенные слова в общем случае не связаны друг с другом): “Ракетносцы просят рубля” (Т, 2002, № 9); <Батальоны просят огня> — название кинофильма; /Сегодня на взлетном поле авиационного производственного объединения стоят несколько замороженных недостроенных ракетноносцев/;

5.3* — *разнесенных слов с сохранением отношения между заменяемыми и заменяющими словами* (ча-

ще всего речь идет об отношениях тождества, антонимии и т.п.): “*Играть* или не *играть*? Вот в чем вопрос!” (И, 2000, № 180); <*Быть* или не *быть*? Вот в чем вопрос!>; /Передозировка компьютерных игр способна оказать негативное влияние на психику не только ребенка, но и взрослых людей/; “*У трезвого всегда нетрезвый виноват*” (АиФ, 2001, № 52); <*У сильного всегда бессильный виноват*>; “*Сколько Лен, сколько Эин*” (Труд-7, 2001, № 44); <*Сколько лет, сколько зим*>; /*Сильная* половина человечества по здравляет прекрасную половину с 8 марта/;

5.4* — *сводящаяся к перестановке двух слов*: “*Пусто место свято не бывает*” (И, 2000, № 78); <*Свято место пусто не бывает*>; /*Есть* такие места работы, где денег вроде бы можно заработать много, но идти туда люди не слишком торопятся/;

- 6* — *блочная замена* (блок из нескольких подряд следующих слов заменяется другим блоком; консервативные элементы прототипа и заголовка в состав блоков (заменяемого и заменяющего) обычно не включаются, иначе любую схему трансформации прототипа в заголовок можно было бы рассматривать как одну блочную замену; блоки могут иметь разную длину; возможны искажения синтаксической структуры прототипа; случаи замены двух соседних слов двумя другими с сохранением синтаксической структуры отнесены к п. 5.1): “*Переход кандидата наук через Альпы*” (И, 2000, № 184); <*Переход Суворова через Альпы*> — название картины Сурикова; /*Десятки тысяч ученых из России работают за рубежом*/ . Здесь имеет место блочная замена типа 1 → 2 (одно слово прототипа заменяется двумя в заголовке); “*Во власть идти — не поле перейти*” (АиФ, 2001, № 50); <*Жизнь прожить — не поле перейти*>; Здесь имеет место блочная замена типа 2 → 3 с изменением синтаксической структуры;
- 7* — *блочная делеция* (устранение блока из нескольких слов):

- 7.1* — *устраняется префиксная часть* прототипа: “Табачок взр^ь” (И, 2000, № 239); <Дружба дружба^й, а табачок взр^ь>; /Все виды легкой атлетики поделили на премьерные и классические .../;
- 7.2* — *устраняется суффиксная часть* прототипа: “Служить бы рад... >” (АиФ, 2002, № 5); <Служить бы рад, прислуживаться *тошно*>; /Когда решится вопрос об альтернативной службе в армии?/;
- 7.3.* — *остальные случаи*: “Нормально, Константин” (И, 2000, 121); <Нормально, Григорий? Отлично, Константин!> — из монолога М. Жванецкого; /Тягов снова стал самарским губернатором .../;
- 8* — *блочная вставка*:
 - 8.1* — *префиксная* (в начало прототипа): “Кто ж нам свинью подложил” (Труд-7, 2001, № 40); <Подложить свинью>; /Ветслужба РФ приостанавливает ввоз к нам свинины из Великобритании/;
 - 8.2* — *суффиксная* (в конец прототипа): “Расцвела сирень, черемуха в саду, и мы вспомнили, что такое аллергия” (Т, 2001, № 85); <Расцвела сирень, черемуха в саду>;
 - 8.3* — *другие случаи* (см. последний пример в п. 10*);
 - 9 — *контаминация* (формирование заголовка на основе нескольких (обычно двух) прототипов): “Красному петуху — море по колено” (Т, 2001, № 41); <Пусть *красного петуха*> — первый прототип; <Пьяному *море по колено*> — второй прототип; /Пожар, возникший в сауне, пожарным пришлось заливать в течение часа/;
 - 10* — *комбинированные схемы варьирования*, сводящиеся к сочетанию двух-трех приемов, из числа представленных выше: “Если кто-то кое-где у нас порой” (И, 2000, № 135); <Если кто-то кое-где у нас порой *честно жить*>

не хочет> — строка из песни; /В США судят чиновников, купивших для страны зимнюю Олимпиаду/. Здесь заголовок получается из прототипа с помощью двух операций: 0.4 (изменение значения лица местоимения: у нас → у них) и 7.2 (устранение суффиксной части прототипа — “честно жить не хочет”). Цифровой код, соответствующий данной комбинированной схеме, имеет вид 10 (0.4; 7.2). Другой пример: “В спорах о гимне рождается мелодия” (И, 2000, № 202); <В споре рождается истина.>; /Читатели “Известий” не обсуждают текст будущего гимна — они выбирают мелодию.../. Здесь использованы три операции, идентифицируемые по числовому коду 10 (0.2; 1.5; 8.3);

- 11* — *сложные случаи* (варьируются сразу несколько компонентов прототипа; разрушается синтаксическая структура; наличие скрытого цитирования угадывается порой лишь по одному–двум характерным словам или фрагментам синтаксической структуры; формальное число операций, переводящих прототип в заголовок, достаточно велико (как правило, больше трех)): “Они думают, что их паровоз летит вперед. Но они уже тени истории” (ЛиФ, 2001, № 38); <Наш паровоз вперед лети> . Следует заметить, что четкой грани между 10 и 11 не существует. Некоторые комбинированные схемы варьирования со значением редакционного расстояния между прототипом и заголовком, равным 3, также обладают отличительными признаками, перечисленными выше.

Отметим, что приведенная классификация покрывает всю анализировавшуюся подборку, хотя не гарантирует в общем случае, что зафиксированный набор редакционных операций всегда приводит к минимальному значению редакционного расстояния между прототипом и заголовком. Например, можно было бы внести операцию *переноса (транспозиции)* слова из одной позиции в другую, что эквивалентно двум операциям из числа имеющихся (устранение слова в одной позиции плюс вставка в другой). Однако этого не сделано, поскольку транспозиция встречается

довольно редко и, как правило, не в чистом виде, т.е. в сочетании с другими операциями; кроме того, в случае минимального по расстоянию переноса функцию транспозиции берет на себя операция перестановки соседних слов.

Предложенный набор редакционных операций не всегда приводит к единственному варианту перевода прототипа в заголовок. Например, заголовок "Маслом дела не испортишь" (Труд-7, 2001, № 40) может быть получен из прототипа <Кашу маслом не испортишь> двумя способами: комбинированием вставки и делеции (10 (2.2; 3.1)) или перестановки и замены (10 (5.4; 1.5)). Такие случаи не слишком часты. Предпочтение отдается варианту с минимальным числом операций (так, двойная замена (код 5.1, к примеру) будет обладать приоритетом над двумя одиночными заменами). В случае равенства числа операций случайно выбирается любой из вариантов.

3.2. Количественные оценки частоты использования разных схем варьирования прототипов в анализировавшейся подборке приведены в табл. 2.

Отметим наиболее существенные особенности представленных данных.

1. Доля нетрансформируемых прототипов (код 0.1), используемых в качестве заголовка, довольно значительна (порядка 20%). Это несколько диссонирует с утверждением из [2] о том, что "прецедентные тексты (или прототипы в нашей терминологии) почти всегда выступают в деформированном виде". Следует добавить, что формально нетрансформированными являются и заголовки с кодом 1.4, но меняется их смысл. Блочные делеции (префиксные (7.1) и суффиксные (7.2)) так же зачастую оставляют нетрансформированной наиболее значимую часть прототипа, имеющую самостоятельное значение.

2. Доля преобразований, описываемых кодами 0.2 ÷ 0.4, формально невелика, но следует учитывать, что многие из них встречаются в сочетании с другими преобразованиями, поэтому отнесены к комбинированным схемам варьирования (код 10). Например, преобразование типа "изменение падежа" даже не включено в столбец с кодом 0, поскольку практически всегда встречается лишь в сочетании с заменой слова.

Т а б л и ц а 2

Относительные частоты встречаемости ($F(\%)$) различных схем варьирования

Схема варьир.-я (код)	$F(\%)$	Схема варьир.-я (код)	$F(\%)$	Схема варьир.-я (код)	$F(\%)$	Схема варьир.-я (код)	$F(\%)$
0	22,9	1	28,8	2	1,8	3	2
0.1*	20,2	1.1	1,4	2.1	0,7	3.1	0,6
0.2	1,1	1.2	7,2	2.2	0,8	3.2	0,5
0.3	0,7	1.3	6,7	2.3	0,3	3.3	0,9
0.4	0,9	1.4	1,6	6*	3,6	8*	1,3
4	3,6	1.5	11,9	7*	9,6	8.1*	0,5
4.1	1,3	5*	5,7	7.1*	2,7	8.2*	0,7
4.2	0,7	5.1*	2,6	7.2*	6,3	8.3*	0,1
4.3*	0,3	5.2*	1,3	7.3*	0,6	9	1
4.4	1,3	5.3*	1,4	11*	0,7	10*	18,8
		5.4*	0,4				

3. Замена одного слова в прототипе (код 1) является наиболее часто используемой операцией (28,8%). Детализация замен показывает, что на первом месте по частоте использования стоит замена словом из того же лексико-грамматического класса (11,9%). Этот тип замен нуждается в последующей более тонкой дифференциации. Довольно высокочастотными являются замены паронимического типа (6,7%) и словами из того же таксономического класса (7,2%).

4. Вставки и удаления одиночных слов представлены довольно слабо (1,8% и 2% соответственно). Почти 15-кратный разрыв в частотах использования одиночных замен по сравнению со вставками и делециями следует признать очень значительным. Интересно отметить, что при варьировании слов путем замены или вставки одного символа в произвольной позиции число допустимых замен (т.е. таких, которые приводят к осмысленному слову) всего лишь в три раза превышает число допустимых вставок [3].

5. Совсем иначе выглядит ситуация с блочными делециями (код 7) и вставками (код 8). Они используются значительно чаще, чем вставки и удаления одиночных слов, к тому же наблюдается двойная асимметрия. А именно: частота блочных делеций примерно в 7 раз превышает частоту вставок, в свою очередь, частота суффиксных делеций (код 7.2) более чем в два раза превышает частоту префиксных делеций.

Высокая частота использования блочных делеций (6,3%) может быть обусловлена несколькими причинами: а) стремлением уменьшить длину заголовка (он должен быть лаконичным); б) необходимостью устранения той части прототипа, которая не имеет отношения к содержанию статьи, для которой подбирается заголовок; в) этическими соображениями, заставляющими устранить именно ту часть прототипа, которая, хоть и в иносказательной, но излишне резкой и/или насмешливой форме содержит негативную оценку описываемых фактов, событий и связанных с ними лиц. Цель, тем не менее, достигается именно благодаря сочетанию корректной формы с возможностью восстановления недосказанного по той части прототипа, что приведена в заголовке: "Хоть горшком назови" (И, 2000, 107); <Хоть горшком

назови, только в печку не ставь>; /Правительство, например, разместило себя на gov.ru .../. Вариант "в" применения блочных делеций является наиболее интересным, и поэтому встречается гораздо чаще, чем "а" и "б".

Асимметрия в частотах использования префиксных и суффиксных блочных делеций объясняется, по-видимому, тем, что обычно именно суффиксная часть прототипа является резюмирующей (особенно в поговорках и пословицах), а поскольку вариант "в" является преобладающим по частоте, это приводит к существенному нарушению баланса между префиксными и суффиксными делециями.

6. Довольно высокая частота использования согласованных двойных замен (код 5, $F = 5,7\%$) обосновывает целесообразность выделения их в отдельный подкласс. Отметим, что преобразования 5.1 и 5.2 преследуют цель сохранения синтаксической структуры прототипа, тогда как преобразование 5.3, также не искажая синтаксическую структуру прототипа, сохраняет еще и тип отношения между разнесенными словами (антонимия и т.п.).

7. Редки, хотя и интересны, случаи контаминации (код 9, $F = 1\%$) и перестановки слов (код 5.4, $F = 0,4\%$). Часть перестановок скрыта в 10, поскольку фигурирует в сочетании с другими преобразованиями. Некоторые перестановки носят "технический характер" (например, при двойной замене иногда удобно поменять слова местами, чтобы сохранить ритмическую структуру прототипа). На семантику это существенно не влияет. Для контаминации характерно часто наличие общего слова в используемых прототипах, по которому и происходит их "склеивание".

8. Группа, объединенная кодом 10 (комбинированные схемы варьирования), стоит в частотном упорядочении на втором месте ($F = 13,8\%$). Она нуждается в отдельной детализации (но уже на большей статистике) с целью выявления устойчиво повторяющихся комбинаций преобразований. Но уже на данном этапе можно отметить, что в большинстве случаев речь идет о комбинации из двух преобразований. Бо́льшее число одновременно используемых преобразований, по-видимому, уже делает прототип неузнаваемым, т.е. исчезает сам смысл цитирования,

подразумевающей в той или иной форме иносказание в сочетании с остроумной игрой слов.

Заключение

На материале подборки, составленной из нескольких тысяч газетных заголовков, построенных по принципу варьирования хорошо известного (находящегося "на слуху") прототипа, получены качественные и (впервые) количественные (частотные) характеристики прототипов, из источников и схем варьирования. Некоторые схемы (например, блочные суффиксные делеции, двойные замены разнесенных слов с сохранением отношений между ними и др.) выделены впервые. Значимость их подтверждена достаточно высокой частотой встречаемости.

Наиболее интересными в содержательном плане являются схемы варьирования, обыгрывающие многозначность слов, фонетическое сходство (замены паронимического типа), использующие префиксные и суффиксные блочные делеции, перевод утвердительных форм в отрицательные и наоборот, контаминации и ряд других.

Результаты работы представляют интерес в плане сопоставления способов трансформации языковых единиц на разных иерархических уровнях, реализации эффективных схем информационного поиска (варьирование запроса), а также выработки подходов к автоматизации отдельных схем варьирования, что в перспективе позволило бы создать компьютерную систему поддержки поиска прототипа для заголовка и трансформации его в нужном направлении.

Л и т е р а т у р а

1. ЗЕМСКАЯ Е.А. Цитация и виды ее трансформации в заголовках современных газет // Поэтика. Стилистика. Язык и культура. Памяти Татьяны Григорьевны Винокур. — М.: Наука, 1996. — С. 157–168.

2. КНЯЗЕВ Ю.П. Инвариант и варианты: пути преобразования прецедентных текстов в газетных заголовках // Обработка текста и когнитивные технологии. — Казань: Отечество, 2001. — С. 73–81.

3. ГУСЕВ В.Д., САЛОМАТИНА Н.В. Количественные исследования вариативности языковых единиц // Труды международной научно-практической конференции KDS – 2001 “Знание – Диалог – Решение”. Т. 1. — С.-Петербург: “Лань”, 2001. — С. 186–193.
4. КОСТОМАРОВ В.Г. Русский язык на газетной полосе. — М.: МГУ, 1971.
5. Фразеологический словарь русского языка/ Под ред. Молоткова. Изд. 4-е. — М.: Русский язык, 1986. — 543 с.
6. Мелерович А.М., Мокиенко В.М. Фразеологизмы в русской речи. Словарь. — М.: Русские словари, 1997. — 864 с.
7. Wagner R.A. Fisher M. J. The string – to – string correction problem // J. ACM: — 1974. — Vol. 21, № 1. — P. 168–173.
8. ГУСЕВ В.Д., САЛОМАТИНА Н.В. Электронный словарь паронимов: версия 2 // НТИ, сер.2: Информационные процессы и системы. — 2001. № 7. — С. 28–33.

Поступила в редакцию
15 января 2003 года