

ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ**Н. Г. Швец****СЕМАНТИЧЕСКАЯ МОДЕЛЬ ВЗАИМОСВЯЗИ ВЕРБАЛЬНЫХ И НЕВЕРБАЛЬНЫХ СОСТАВЛЯЮЩИХ КРЕОЛИЗОВАННОГО ТЕКСТА**

В статье решаются задачи построения семантической модели взаимосвязи вербальных и невербальных составляющих рекламного объявления. Предлагается верифицированная методика установления корреляций между вербальным и невербальным (изображением) компонентами креолизованного текста, основанная на определении их близости через максимальное число совпадений опорных слов вербальной составляющей и дескрипторов, описывающих изображение. Продемонстрирована эффективность применения тезаурусного подхода к формализации содержания изображения.

Исследование механизмов порождения и восприятия смысла креолизованных текстов представляет особый интерес в связи со стремительным развитием современных информационных технологий, направленных на решение проблем автоматизированной обработки семиотически неоднородных массивов информации.

Текст и изображение по физической структуре представляют собой два разных типа информации: прерывный (знаковый) и непрерывный (аналоговый) соответственно. Принципы восприятия человеком знаковой и аналоговой информации во многом отличаются. На расшифровку текста требуются время и усилия, а аналоговая информация (образ), напротив, быстро «узнается» на уровне сознательного внимания и напрямую (без декодирования) воспринимается подсознанием. При этом восприятие аналоговой информации не носит четко обусловленного характера и не подчиняется определенно заданным правилам [1, с. 16].

Начало научному осмыслению креолизованных текстов было положено в работах по семиотике, в которых изображения рассматривались как особая знаковая система. Объектом исследования семиотики текста, помимо изучения невербальных средств письменного текста, является и «весь этот текст как единое целое, образуемое из синтеза вербальных и невербальных знаков, а также характер взаимоотношений и взаимодействий этих знаков в оптимальном изображении и передаче информации» [2, с. 39].

К креолизованным текстам относится и печатное рекламное объявление (РО), в котором вербальные и иконические элементы образуют одно визуальное, структурное, смысловое и функциональное целое, нацеленное на комплексное прагматическое воздействие на адресата. Исследователи отмечают, что эффективное сочетание в рекламе вербального знака и иллюстрации возможно лишь в результате длительной работы по разработке композиции РО. И проблема такого эффективного сочетания остается серьезной проблемой для современной лингвистической науки.

Цель настоящего исследования – найти оптимальный способ взаимосвязи вербальной составляющей РО и иллюстрации этого РО и представить ее в виде алгоритма.

Актуальность предлагаемого исследования обусловлена, во-первых, неослабевающим интересом к анализу самых разных аспектов рекламного дискурса, который во многом стимулирует современные исследования в области анализа дискурса в целом; во-вторых, значимостью исследований механизмов порождения и восприятия смысла креолизованных текстов, в которых в рамках единого сообщения сплетены элементы различных семиотических кодов; в-третьих – необходимостью разработки способов формального представления содержания таких текстов.

Материалом исследования послужили печатные РО на русском языке, взятые из различных журналов за 2000–2019 годы. В качестве источников использовались русскоязычные версии журналов “Cosmopolitan”, “Woman”, “ELLE”, “Burda”, журналы «Натали», «Домашний очаг», «Лиза» и др. Всего для анализа было отобрано 1 000 РО по теме «Косметика и парфюмерия», содержащих изображение и достаточно большую вербальную часть (длиной от трех и более абзацев).

Гипотезой исследования стало предположение о том, что, представляя содержание текста РО в виде определенного набора главных и второстепенных опорных слов, а содержание иллюстрации РО – в виде многоуровневого комплекса дескрипторов, можно определить близость этих содержаний через максимальное число совпадений опорных слов и дескрипторов с учетом их степени важности.

Для того, чтобы проверить данную гипотезу, нужно построить семантическую модель взаимосвязи вербальных и невербальных составляющих РО, решив следующие задачи:

- 1) выделить основные составляющие РО;
- 2) найти наиболее оптимальный для достижения поставленной цели формальный способ представления основного содержания текста РО;
- 3) выявить наиболее эффективный для достижения поставленной цели способ формального представления содержания иллюстрации РО;
- 4) разработать лингвистическую базу знаний, содержащую формальные описания текстов и изображений исследуемых РО;
- 5) разработать формальную модель, позволяющую автоматически выбирать для заданного текста РО возможную по содержанию иллюстрацию;
- 6) проверить эффективность работы созданной модели с помощью компьютерной программы.

Конкретизируя в целях нашего исследования понятие *рекламное объявление*, будем называть им семиотически неоднородный текст, содержащий вербальный (словесный) компонент (непосредственно рекламный текст) и визуальный (невербальный) компонент (изображение), представленный в письменной форме, заранее подготовленный, обладающий автономностью, направленный на донесение до адресата определенной информации с целью привлечения внимания к тому или иному виду товара.

С точки зрения композиционной организации в тексте РО можно выделить следующие основные логико-семантические составляющие:

1) привлекающий компонент – заголовок, который выражает основную идею текста;

2) разъясняющая часть – основной рекламный текст (или информационный блок), который развивает положения, выдвигаемые в заголовке;

3) управляющий компонент – слоган и эхо-фразу, которые задают алгоритм действий адресата.

Невербальный компонент РО могут составлять: фотографии, рисунки, фирменный знак, национальная символика, схема, чертеж, диаграмма, шрифт, цветовое оформление, декоративные элементы, а также сегментация текста, пробелы, знаки препинания, заглавные буквы, искажения орфографии, т.е. все то, что придает тексту прежде всего зрительную выразительность. В настоящем исследовании во внимание принимается лишь одна невербальная составляющая РО – иллюстрация, которая может быть представлена в виде фотографии, рисунка, схемы.

В большинстве случаев в изображениях РО используется конкретная внешняя наглядность, представляющая определенные предметы рекламы, людей, ситуации (например, демонстрация товара в процессе его использования, иллюстрирование характеристик товара и т.д.) [3].

Однако изображение в РО может иллюстрировать передаваемый смысл не только прямо, но и косвенно, например, когда изображенные предметы или ситуации не имеют «портретного» сходства с передаваемым содержанием, а являются его образным эквивалентом. Кроме того, в рекламе могут быть использованы рисунки с намеком на какое-то действие, т.е. с неразвернутой ситуацией. В таких случаях мы имеем дело с внутренней наглядностью, которая представляет собой сумму индивидуального опыта реципиента и в силу этого его персонального осмысления изображенной ситуации. Внутренняя наглядность основывается на образе представления и домыслах во всем их богатстве и разнообразии, так как неразвернутость ситуации является предпосылкой для самого широкого использования домысла, фантазии.

Таким образом, можно выделить РО с жестким (прямолинейным) смыслом, в которых используется конкретная внешняя наглядность изображения и РО с многозначным (неясным, расплывчатым) смыслом, в которых приводятся изображения с неразвернутой ситуацией.

Для решения следующей задачи нашего исследования необходимо, прежде всего, определить понятие *основное содержание текста*. Под ним понимается совокупность знаменательных слов, несущих основную семантическую нагрузку в тексте, формирующих в сознании адресата то же представление об описываемом фрагменте ситуации, которое вложил в текст автор [4].

Детальный анализ различных современных подходов к способам формального представления основного содержания текста позволил нам сделать вывод, что формальное представление основного содержания текста является эффективным и легко реализуемым на компьютере при учете общей частоты употребления знаменательного слова и всех его словарных и контекстуальных замен в тексте и его абзацах.

Вследствие того, что определение основного содержания текста путем использования одной группы методов не обеспечивает удовлетворительного качества результатов, такие методы, как правило, дополняют друг друга. Методика комбинированного подхода к представлению основного содержания текста, сочетающая в себе элементы статистического и позиционного методов извлечения основной смысловой информации, позволяет существенно повысить корректность автоматического извлечения ключевых слов.

Тема текста (или его основное содержание) может быть представлена в виде таблицы основного содержания текста, основными компонентами которой являются опорные слова, разделенные в соответствии с предметными свойствами своих референтов на группы слов (слова-объекты, слова-признаки, слова-действия и прочие слова). В результате были получены такие таблицы для текстов 1000 РО.

Таким образом, в нашем исследовании основное содержание анализируемых текстов РО представлено формальным образом с помощью набора главных и второстепенных опорных слов, дополненных ключевыми словами из заголовков РО. Эти слова несут основную семантическую нагрузку в тексте, обеспечивая становление и развитие темы. Преимущество выбранной нами методики состоит в возможности классифицировать слова конкретного текста в зависимости от степени их важности для семантической структуры текста по нескольким группам, а также в гибкой применимости данной методики к текстам РО с разным количеством абзацев.

Среди возможных способов формального представления иллюстрации нами используется тезаурусное представление изображения как способ, позволяющий учесть не только доминирующие, но и второстепенные черты изображения. Оно позволяет описать иллюстрацию в виде некоторого набора слов (дескрипторов) и отношений между этими словами в рамках некоторой предметной области.

В настоящем исследовании разработана методика тезаурусного описания изображений РО по теме «Косметика и парфюмерия», включающая следующие этапы [5].

1. Выявление языкового, денотативного и коннотативного содержаний для каждого изображения исследуемых РО.

2. Определение дескрипторов, образующих первый уровень тезауруса.

3. Выделение следующих типов отношений между дескрипторами: ЦЕЛОЕ – ЧАСТЬ, СОМАТЕМА – ПРИЗНАК, СОМАТЕМА – ДЕЙСТВИЕ, ОБЪЕКТ – ПРИЗНАК, ОБЪЕКТ – ДЕЙСТВИЕ, АССОЦИАЦИЯ.

4. Систематизация отношений типа СОМАТЕМА – ПРИЗНАК по ряду параметров (цвету форме и т.п). Такие параметры были выделены для описания следующих соматем: «Женщина», «Волосы», «Глаза», «Губы», «Зубы», «Кожа», «Лицо», «Нос», «Ресницы», «Тело (Фигура)».

5. Систематизация отношений типа ОБЪЕКТ – ПРИЗНАК по следующим параметрам: «форма», «цвет», «содержимое», «название», «фирма-изготовитель».

В результате использования тезаурусного подхода были построены многоуровневые тезаурусные описания изображений 1 000 РО в виде некоторого набора слов (дескрипторов) и отношений между этими словами. Пример одного из таких описаний представлен в табл. 1.

Таблица 1

Пример тезаурусного представления иллюстрации РО

Дескриптор (уровень тезауруса)	Мероним (уровень тезауруса)	Признак (уровень тезауруса)	Действие (уровень тезауруса) (ассоциация)
Женщина (1)	Голова (2) Руки (2)		
Фрукты (1)	Апельсины (2) Лимоны (2) Черника (2) Витамины (2)		
Баночка (1)	Крем (2)	Белая (2) Цилиндрическая (2) Раскрытая (2)	
Голова (2)	Лицо (3) Волосы (3)		
Руки (2)	Персик (3)		
Витамины (2)			Воздействуют на кожу (3) (эластичность и молодость кожи)
Крем (2)		Synergie – Сияние Витаминов (Сияние Витаминов Synergie) (3) Garnier Paris (Гарнье Париж) (3)	
Лицо (3)	Кожа (4) Улыбка (4)		
Персик (3)	Витамины (4)	Румяное (4) Свежее (4)	Соприкасается с кожей лица (5) (свежесть лица; лицо румяное, как персик)
Кожа (4)		Молодая (5) Свежая (5) Эластичная (5) Нежная (5) Увлажненная (5) Красивая (5) Румяная (5) Сияющая (5)	

Следующая задача исследования была связана с поиском наиболее эффективного формального способа взаимосвязи содержания вербальной составляющей и иллюстрации РО. Для этого был предложен принцип максимального совпадения семантических вех вербального компонента (опорных слов) и изображения (дескрипторов) на разных уровнях семантической значимости этих единиц. Этот принцип и стал основой формальной модели семантической взаимосвязи двух исследуемых частей РО.

Под формальной моделью взаимосвязи текста и иллюстрации РО в настоящем исследовании принято представление этой связи на лексико-семантическом уровне в виде алгоритма [6].

Для реализации алгоритма в виде компьютерной программы была создана база знаний, состоящая из следующих компонентов:

1) текстов РО, отобранных для компьютерного эксперимента, относящихся к трем предметным областям «Крем для лица», «Шампунь», «Краска для волос»;

2) изображений к текстам исследуемых РО;

3) таблиц основного содержания исследуемых рекламных текстов, заданных в виде перечня главных и второстепенных опорных слов каждого такого текста;

4) формальных представлений изображений исследуемых РО, заданных в виде тезаурусных описаний этих изображений.

Построенный алгоритм был реализован в виде компьютерной программы, написанной на языке C#.

В процессе тестирования работы программы были получены результаты, представленные в табл. 2.

Таблица 2

Результаты тестирования компьютерной программы

Предметная область	Кол-во проанализированных текстов РО	Кол-во безошибочно выбранных иллюстраций, %
Шампунь	74	85
Крем для лица	66	88
Краска для волос	60	86
Всего	200	86

Анализ результатов работы программы свидетельствует о достаточно высокой эффективности разработанной формальной модели: количество случаев, когда пара «текст – изображение» была точно восстановлена, составило 86 %.

В предложенной нами схеме есть два важных субъективных фактора: первый – это тот, кто создает изображение РО (назовем его Иллюстратор), и тот, кто описывает изображение с помощью тезауруса (назовем его Эксперт).

Если и Иллюстратор, и Эксперт профессионально выполняют свои функции (хорошо обучены), то пара «текст – изображение» восстанавливается. Именно это и показал эксперимент. Однако возможны погрешности при сравнении формальных представлений текста и изображения РО, которые составили 14 %.

Если Иллюстратор плохо обучен, то содержание иллюстрации РО плохо соответствует содержанию текста. В связи с этим пара «текст – изображение» не восстановится, даже если Эксперт справится со своей задачей хорошо.

Если Иллюстратор обучен хорошо, то содержание иллюстрации соответствует содержанию текста РО. И тогда все зависит от компетентности Эксперта. Если тезаурусное описание будет хорошим, то пара «текст – изображение» восстановится. Если Эксперт плохо обучен, то возникает большее количество ошибок в выборе иллюстрации. Именно об этом говорят результаты второго проведенного нами эксперимента (табл. 3), когда в роли Экспертов выступали студенты 5 курса факультетов английского, немецкого, романского языков и межкультурных коммуникаций МГЛУ, которые описали с помощью тезаурусного подхода 200 изображений РО, используя разработанную нами методику.

Таблица 3

Результаты второго эксперимента

Предметная область	Кол-во проанализированных текстов РО	Кол-во безошибочно выбранных иллюстраций, %
Шампунь	74	75
Крем для лица	66	77
Краска для волос	60	74
Всего	200	75

Таким образом, экспериментально-опытная проверка предлагаемой модели показала, что существует взаимосвязь вербального текста и иллюстрации РО продемонстрировала эффективность применения тезаурусного подхода к формализации содержания изображения.

ЛИТЕРАТУРА

1. Гребенкин, Ю. Ю. Психотехнологии в рекламе / Ю. Ю. Гребенкин. – Новосибирск: Издат. дом «РИФ-плюс», 2000. – 214 с.
2. Плотников, Б. А. Семиотика текста: Параграфемика: учеб. пособие / Б. А. Плотников. – Минск.: Высш. шк., 1992. – 190 с.
3. Ромат, Е. Реклама / Е. Ромат. – СПб.: Питер, 2002. – 544 с.
4. Швец, Н. Г. Определение основного содержания текстов рекламных объявлений / Н. Г. Швец // Материалы ежегод. науч. конф. преподавателей и аспирантов ун-та, 16–17 апр. 2002 г.: в 3 ч. / Мин. гос. лингвист. ун-т; отв. редактор Н. П. Баранова. – Минск, 2003. – Ч. 2. – С. 82–87.

5. Швец, Н. Г. Формальный способ представления изображений рекламных объявлений в виде тезауруса / Н. Г. Швец // Вестн. МГЛУ. Сер. 1, Филология. – 2019. – № 2. – С. 161–171.

6. Швец, Н. Г. Информационная система семантического поиска изображения к тексту рекламного объявления / Н. Г. Швец // Системный анализ и прикладная информатика. – 2019. – № 4. – С. 62–68.

The article addresses the problem of constructing a semantic model of the relationship between verbal and non-verbal components of an advertisement. A verified method is proposed for establishing correlations between the verbal and non-verbal (image) components of a creolized text, based on determining their similarity through the maximum number of matches of the reference words of the verbal component and descriptors describing the image. The effectiveness of applying the thesaurus approach to formalizing image content is demonstrated.

Поступила в редакцию 09.06.17