

81
Д. 386

МИНСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ ИНСТИТУТ
ИНОСТРАННЫХ ЯЗЫКОВ

На правах рукописи

ДЕТСКАЯ Раиса Владимировна

СТРУКТУРНО-СЕМАНТИЧЕСКИЙ И КОММУНИКАТИВНЫЙ
СТАТУС ЗАГЛОВОКОВ РЕФЕРАТОВ К ПАТЕНТАМ В СИСТЕМЕ
АВТОМАТИЧЕСКОЙ ПЕРЕРАБОТКИ ТЕКСТА

Специальность 10.02.19 - теория языкознания

А в т о р е ф е р а т
диссертации на соискание ученой степени
кандидата филологических наук

Минск - 1991

Работа выполнена в Минском государственном педагогическом институте иностранных языков.

Научный руководитель: кандидат филологических наук,
доцент Т. П. КАРПИЛОВИЧ

Официальные оппоненты: доктор филологических наук,
профессор В. А. ЧИБАКОВСКИЙ
кандидат филологических наук,
и.о. профессора Т. С. ГЛУШАК

Ведущая организация: Тартуский университет

Защита состоится "22" сентября 1991 г. в "14"
часов на заседании специализированного совета Д.056.06.01 по
присуждению ученой степени доктора филологических наук в
Минском государственном педагогическом институте иностранных
языков по адресу: 220034, г. Минск, ул. Захарова, 21.

С диссертацией можно ознакомиться в библиотеке Минского
государственного педагогического института иностранных языков.

Автореферат разослан "20" сентября 1991 г.

Ученый секретарь
специализированного совета
доктор филологических наук,
профессор

А. Мещ

Н. В. ШЕРЕМЕТА

Реферируемая диссертация представляет собой опыт решения с позиций ведущих направлений современного языкознания — коммуникативной лингвистики, языковой номинации, денотативной семантики — общей проблематики заголовка в его связи с текстом и конкретного исследования информативно-коммуникативной значимости немецких заголовков рефератов к патентам по складскому хозяйству.

Интерес к вопросам, составляющим проблематику заголовков научно-технических текстов, стал заметно возрастать. Это объясняется как расширением исследований в области лингвистики текста, так и новыми возможностями изучения заголовка, связанными с достижениями в области автоматической переработки текста. Растущий спрос на быстрое и массовое информационное обслуживание учреждений и предприятий обуславливает особый интерес к машинной обработке заголовков научных статей, поскольку в них содержится ключ к содержанию текста.

Как известно, заглавия научных трудов обозначают, как правило, область исследования, общую направленность или тему работы, дифференциацию смыслового объема. Эта особенность их семантико-функциональной стороны находится в прямой связи с сутью обработки информации, при которой научные работы тематически классифицируются по содержанию заголовков.

К настоящему времени проведено значительное количество исследований, посвященных анализу заголовка как компонента текста^I. Однако, при всей актуальности рассматриваемых в них проб-

^I Шамелашвили М.А. Функционально-стилистические и лексико-грамматические особенности газетных заголовков: Автореф. дис. ... канд. филол. наук. — М., 1982. — 24 с.; Пшеничная Л.Э. Модель раскрытия содержания заголовка средствами полного текста научного документа // Создание автоматизированных систем информационного обеспечения научных исследований. — Киев, 1986.

дем ряд вопросов организации заголовков остается все еще нерешенными. Так, в частности, не получил достаточного освещения вопрос о структурно-семантических свойствах заголовков научно-технических текстов, их связи с общим содержанием текста.

В данной работе выбор объекта исследования как раз и продиктован указанной причиной, точнее отсутствием общепринятого взгляда на грамматическую природу заголовка, его структурный и семантико-синтаксический статус, специфику связи с текстом определенного жанра научно-технической литературы. Изучение этих аспектов неизбежно выводит в проблематику вопросов, находящихся в центре внимания современных лингвистов: моделирование предложений; коммуникативная организация предложения и текста; соотношение предложения, высказывания и текста и др.

Актуальность выполненного исследования обусловливается необходимостью раскрытия роли заголовков в коммуникативной структуре научно-технических текстов определенного жанра, потребностями разработки алгоритмов анализа речевых единиц как основы систем автоматической переработки текста.

Материалом исследования явились 5600 заголовков рефератов к патентам на немецком языке по тематике "Организация складского хозяйства и его материально-техническая база" и сами тексты рефератов данной предметной области. Компрессированный и стандартный характер заголовков такого рода научно-технических документов позволяет представить в виде определенного конечного списка структурные и коммуникативные типы исследуемых заголовков.

С.36-45; Чижиковский В.А. Семантико-коммуникативные аспекты автоматической переработки заголовка научно-технического текста. Автореф. дис. ... доктора филол. наук. - Л., 1989. - 31 с.

Основную цель диссертации составляют разработка на основе результатов лингвистического анализа отобранного материала алгоритма тема-рематического членения заголовков и создание экспериментальной системы немецко-русского машинного перевода (МП) заголовков рефератов к патентам.

Поставленная цель предполагает решение следующих конкретных задач:

- 1) построение структурно-синтаксических моделей немецких заголовков рефератов к патентам;
- 2) анализ семантико-синтаксической структуры заголовков;
- 3) изучение роли заголовка в организации коммуникативной структуры текста реферата к патенту;
- 4) выявление коммуникативных типов заголовков рефератов к патентам и разработка алгоритма их автоматического определения;
- 5) установление структурных соответствий немецких заголовков в русском языке;
- 6) разработка алгоритма и программы немецко-русского МП заголовков рефератов к патентам.

В работе использованы следующие методы исследования: сопоставительный, контекстуальный, статистический, семантико-синтаксический, логико-семантический анализ, а также метод моделирования и некоторые частные процедуры.

Научная новизна проведенного исследования заключается в том, что информативно-коммуникативная значимость заголовка реферата к патенту установлена и обоснована в непосредственной связи с озаглавленным текстом. На базе анализа количественных данных и характера повторов лексических номинаций заголовка в тексте реферата выявлена роль компонентов семантической структуры заголовка в тема-рематической организации всего реферата.

Выявлены также коммуникативные типы заголовков, обусловленные различной позицией заголовочных предикатов, входящих в рематический компонент текста реферата. С учетом коммуникативных типов заголовков и их маркеров — базовых синтаксических моделей — разработан алгоритм тема-рематического членения заголовков, который может быть использован в автоматизированных системах информационного поиска и распознавания смысла научно-технического текста.

На основе установленного набора базовых моделей немецких заголовков, их структурных соответствий в русском языке и диагностирующих признаков создан алгоритм немецко-русского МП заголовков, реализованный на персональной ЭВМ.

Теоретическая значимость диссертации связана с тем, что полученные результаты могут способствовать дальнейшей разработке теоретических положений о семантической и коммуникативной структуре заголовка как неотъемлемого компонента текста.

Практическая значимость работы заключается в возможности применения созданного алгоритма тема-рематического членения немецких заголовков рефератов к патентам в автоматизированных системах информационного поиска и распознавания смысла научно-технических документов, а также при создании компьютерных обучающих программ. Разработанная система МП заголовков реализована в экспериментальном режиме и может быть использована для оперативного получения сигнальной информации о новых изобретениях.

На защиту выносятся следующие положения диссертации:

1. Заголовок является предложением особого рода, коммуникативное назначение которого состоит в передаче основного содержания текста в краткой и сжатой форме.
2. Занимая одну из "сильных" позиций текста реферата к патенту, заголовок конденсирует текстовую информацию, которая

уточняется и конкретизируется в реферате.

3. Коммуникативность заголовка как семантического репрезентанта текста реферата к патенту определяется его свойством указывать на тематический и рематический компоненты текста реферата.

4. Основные составляющие семантической структуры заголовка (объект описания и заголовочные предикаты) принимают различное участие в организации коммуникативной структуры текста реферата. Объект описания, называя главную тему реферата, обеспечивает его семантическое единство и целостность. Заголовочные предикаты, представляющие характеристики объекта описания, входят в рематическую часть реферата, которая охватывает практически весь его текст, описывающий суть изобретения.

Апробация работы проводилась на Международном семинаре по машинному переводу (Тбилиси, 1989); на Всесоюзных конференциях: "Переработка текста методами инженерной лингвистики" (Минск, 1982), "Использование ЭВМ в научной и учебной работе гуманитарных вузов" (Минск, 1990), "Актуальные проблемы компьютерной лингвистики" (Тарту, 1990); на межвузовской научной конференции "Проблемы компьютеризации процесса преподавания иностранных языков в педагогических институтах" (Ульяновск, 1986); во Всесоюзной школе-семинаре "Бионика интеллекта" (Харьков, 1987); на областной конференции преподавателей кафедр иностранных языков "Лексикографические методы в обучении и научных исследованиях" (Иваново, 1988); на итоговых конференциях Минского ЛПИИ (1980-1990 гг.).

Структура диссертации определяется целью и задачами исследования. Композиционно диссертация состоит из введения, трех глав, заключения, списка использованных источников и пяти приложений.

Во введении дается обоснование актуальности темы, опреде-

ляются цель и задачи, новизна, теоретическая и практическая значимость работы, описывается материал исследования, излагаются основные положения, выносимые на защиту.

В первой главе теоретически обосновываются положения, на основании которых проводится лингвистический анализ, ориентированный на обработку заголовков рефератов к патентам по складскому хозяйству. В ней получает трактовку семантическая природа заголовка, дается функционально-коммуникативная характеристика заголовков научно-технических документов, определяется лингвистический статус заголовка.

Во второй главе проводится структурно-семантический анализ заголовков рефератов к патентам по складскому хозяйству, устанавливается роль заголовка в коммуникативной структуре текста реферата на основе анализа характера и способов повторов лексики заголовков в рефератах, анализируются кореферентные связи лексики заголовка, представляющей основные компоненты его семантической структуры, и текста реферата к патенту, определяются коммуникативные типы заголовков, разрабатывается алгоритм выявления коммуникативных типов заголовков рефератов к патентам.

В третьей главе описывается процедура формализации основных лингвистических данных, полученных в предыдущих главах: устанавливается структурная типология базовых синтаксических моделей немецких заголовков и их русских соответствий, разрабатывается общий и детальные алгоритмы и программа немецко-русского МП заголовков рефератов к патентам по складскому хозяйству.

В заключении излагаются основные результаты исследования, формулируются выводы и определяются возможные направления дальнейших исследований в области затронутой проблематики.

Приложения содержат конкретный лингвистический материал

для построения системы немецко-русского МП заголовков рефератов к патентам и результаты практической реализации на ЭМ ГВ РС/ХТ.

СОДЕРЖАНИЕ РАБОТЫ

Современный уровень развития лингвистической науки позволяет подойти к изучению структурно-функциональных особенностей заголовков в ширком плане - через сопоставление их роли с общей теорией коммуникации. С таких позиций логично признать настоящим отождествление роли заголовков с функцией только "надписей-ярлыков", лишаящее их коммуникативной значимости, т.е. связи с содержанием озаглавленного текста. При подобном отождествлении не учитывается специфика заголовков, а именно тот факт, что они являются частью текста и обладают двусторонней связью как с текстом, так и с непосредственными объектами действительности, отражаемыми в этом тексте.

Заголовку как компоненту текста присущи два основных признака: номинативный характер, с которым связана способность заголовка к предикации, а значит к сообщению чего-то нового о данном явлении, и информативность, заключающаяся в способности заголовка передавать ядро содержания озаглавленного текста. Оба признака позволяют считать, что статус заголовка как компонента текста объединяет в себе название и передачу основного содержания озаглавливаемого текста. Тем самым речь идет о несомненно "сильной позиции".

"Сильная позиция" предопределяет активное участие заголовка в текстообразовании и прежде всего в том плане, что он содержит в большинстве случаев указание на тему и ремю текста. В заголовке поэтому как бы конденсируется текстовая информация, развертывание которой с необходимой степенью эксплицитности

определяет сам текст. Информирова определенным образом адресата о характере и содержании предваряемого им текста, заголовок способствует формированию у него и соответствующей коммуникативной установки. Возможностью "предсказания" последующего содержания или основной идеи текста детерминируется коммуникативная роль заголовка – своего рода смыслового узла, который стягивает все стороны текста. Можно сказать об этом и иначе: заголовок кодирует содержание текста, вбирая его в себя в наиболее общем виде. В целом же правомерно утверждать, что заголовок есть неотъемлемый компонент текста, отмеченный той спецификой, что он предваряет текст (или его сегмент) и конденсирует его содержание, являясь одновременно графическим маркером его семантической самостоятельности и структурной завершенности.

Учитывая двойственную природу заголовка – соотносительность с содержанием объективной действительности (через текст) и со своим собственным денотатом (озаглавленным текстом) – можно выделить в характеристике заголовка два аспекта: 1) коммуникативность заголовка как семантического репрезентанта текста и 2) коммуникативность заголовка как самостоятельного высказывания. Эти аспекты подчеркивают двойственный статус заголовка как компонента текста и как виртуального предложения. Последнее, удовлетворяя требованиям грамматической, семантической и логической правильности, не удовлетворяет условиям целенаправленности и ситуативной соотносительности.

Однако, репрезентативная функция заголовка проявляется не только в его стремлении выразить ядро содержания информации, но и в отражении авторской интенции, что влияет на оценку материальной манифестации заголовка: он информирует даже в случае, если заголовок представлен одиночным номинативом. В номинации заголовка отражаются компоненты концептуально-тематической

основы текста, а это означает, что он совмещает акт номинации (именование текста) и процесс предикации.

Таким образом, наиболее существенным моментом текстообразующих свойств заголовка следует считать его участие в формировании и отражении содержания текста, в способности интегрировать отдельные, концептуально-значимые элементы текста, создавая смысловое единство, а также придавать тексту цельнооформленность и выполнять роль того предела, к которому стремится логическое построение текста.

В сфере научно-технической информации заголовок является важной единицей коммуникации, выступая по существу в функции информатора, при этом с ярко выраженной целевой информацией. К ней необходимо привлечь внимание адресата, чтобы заставить его прочесть материал статьи или соответствующего документа и вызвать определенное отношение к описанному объекту. Для этого заголовок должен, в первую очередь, с семантической точки зрения адекватно передавать содержание документа. Тем самым, авторская интенция направлена как раз на то, чтобы отразить в заголовке основные смысловые аспекты содержания озаглавливаемого текста. Фактически адресат получает информацию в два приема: сначала конденсированно с помощью заголовка, а далее развернуто с помощью самой статьи, прочесть которую заголовок его побуждает.

Выше было отмечено, что для заголовка научно-технического документа характерна одновременная передача и уже известной, и новой информации с опорой на известную. В этом проявляется прагматика заголовка: он сам по себе должен играть роль сокращенного реферата, призванного конденсировать все содержание. Из этого и вытекает смысловая прагматическая сила заголовка — его самостоятельное воздействие на читателя. Все изложенное позволяет сформулировать определенные требования к заголовку: он дол-

жен соответствовать содержанию документа и сочетать в себе максимальную информационную емкость, краткость, конкретность, четкость. Эти требования определяют ряд функций заголовка, результат взаимодействия которых соотносится с жанровой разнородностью текстов научно-технических документов. При общей коммуникативной установке различные научно-технические документы (статья, рецензия, доклад, реферат к патенту) выполняют, как известно, разные коммуникативные задачи. Так, основной коммуникативной задачей реферата к патенту, неотъемлемой частью которого является заголовок, служит сообщение новых сведений - научно-технической или технологической информации - потребителю. В ней имплицитно также юридическая защита прав изобретателя на патентное изобретение. Задача решается через взаимодействие двух основных функций заголовка реферата к патенту - номинативной и коммуникативной. Помимо этого, с точки зрения потребителя, заголовок научно-технического документа выполняет, кроме этих двух функций, еще и рекламную, адресную, поисковую. В системах массовой коммуникации заголовок научно-технического документа реализует все эти функции, при том, что всегда доминирует глобальная общекоммуникативная функция - определение установки читателя на восприятие основного содержания текста через типизируемую в отражении заголовком ситуацию.

Проведенный лингвистический анализ показал, что заголовочные конструкции носят в основном именной характер, по структуре представляя собой номинативные (назывные) предложения. Основной единицей номинации в них является существительное, образующее ряд синтаксических моделей в результате препозитивного и постпозитивного распространения именного ядра. В работе они именуется базовыми синтаксическими моделями. Базовые модели - это n-компонентные словосочетания, включающие только полнозначные

слова. Формальными разделителями базовых моделей служат, главным образом, предлоги, союзы, знаки пунктуации, а также отдельные словоформы.

Если исходить из того, что элементы языка обладают определенными сочетательными возможностями, то следует констатировать, что сочетательные особенности различны у элементов, входящих в рассматриваемые базовые синтаксические модели. Имеется в виду, что они различны, с одной стороны, у существительных, с другой стороны — у прилагательных, причастий, местоимений и числительных, что диктует образование или невозможность образования определенных группировок элементов.

Простейшей структурой анализируемых заголовков рефератов к патентам является S (существительное). Например: *Bandförderer* (81,10,45,3014149)^I; *Dosierspender* (86,1,46,334412). Однако, частотность примеров, в которых заголовок реферата к патенту представлен одиночным существительным, невелика (10,3%). В подавляющем большинстве случаев ядерное существительное имеет левые и правые распространители и их комбинации. В зависимости от синтаксического типа распространителей и их позиции по отношению к ядру, выделено 25 структурных вариантов ядерной модели S. Фрагмент таблицы выделенных базовых синтаксических моделей заголовков приведен на стр.12. Структурные варианты ядерной модели могут усложняться за счет однородных членов, вводимых обычно союзами *und* или *oder*. Таким способом может расширяться любой из компонентов выделенных моделей. Например: *Krabbelwagen und Fördersystem* (88,3,51,354298); *Verpackungsbehälter und*

^I В скобках дается источник примера: 1-я цифра указывает год издания, 2-я — номер журнала, 3-я — номер выпуска, 4-я — номер патентуемого изобретения. Все примеры взяты из реферативного журнала "Изобретения стран мира".

Таблица I (фрагмент)

Базовые синтаксические модели немецких заголовков
рефератов к патентам

№ п/п	Структурные формулы	Примеры	Употребитель- ность выделен- ных моделей в исследуемом массиве заго- ловков	
			к-во	%
ОДНОКОМПОНЕНТНЫЕ МОДЕЛИ				
I. S		Bandförderer (8I, IO, 45, 30I4I49)	I436	IO,3
ДВУХКОМПОНЕНТНЫЕ МОДЕЛИ				
2. S _I +S ₂		Deckel einer Haushaltgross- packung (88,5,49,3600465)	444	3,2
3. A+S		Silo für feinteilige Schüttgü- ter (88,3,49,3538782)	3956	28,4
ТРЕХКОМПОНЕНТНЫЕ МОДЕЛИ				
4. S _I +S ₂ +S ₃		Vorrichtung zum Auftrennen einer Umreifung eines Rohstoffballens (88,2,49,363I707I)	284	2
5. S _I +A+S ₂		Reihenanzordnung elektronischer Bausteine (88,5,49,3644367)	828	6
6. A+S _I +S ₂		Vorrichtung zum absatzweisen Abziehen einer Packmittelbahn (86,23,46,235379)	752	5,4
7. A _I +A ₂ +S		Ineinandersteckbarer, transpor- tabler Container mit Vollring- verschluss (83,II,46,3044302)	450	3,2
ЧЕТЫРЕХКОМПОНЕНТНЫЕ МОДЕЛИ				
8. S _I +S ₂ +S ₃ +S ₄		Vorrichtung zur Beeinflussung der Laufrichtung des Förder- mittels eines Bandförderers (85,9,47,33I3IO9)	2I6	I,6
9. S _I +S ₂ +A+S ₃		Anordnung zur Veränderung der Zugkraft eines laufenden Fadens (86,5,47,3403I44)	490	3,5
IO. S _I +A+S ₂ +S ₃		Vorrichtung zum Verlockern der oberen Bogen eines Bogenstapels (8I,2,45,2906873)	2I2	I,5
II. S _I +A _I +A ₂ +S ₂		Vorrichtung zur Führung stehen- der zylindrischer Behälter (86,6,46,225964)	820	5,9
I2. A _I +S _I +A ₂ +S ₂		Verfahren und Vorrichtung zum automatischen Einlegen kasten- förmiger Säcke (84,3,46,320307I)	558	4

Herstellungsverfahren (86,7,46,3407688).

Анализ массива заголовков показывает, что заголовки рефератов к патентам чаще всего представляют собой комбинации нескольких моделей. Например: (S zum S) - *Vorrichtung zum Palettieren* (87,15,50,3517049); (S als S) - *Papptrommel als Haushaltgrosspackung* (88,5,49,3600521).

Сведение всего многообразия синтаксических структур немецких заголовков рефератов к патентам к 25 базовым моделям, представляющее их в обозримом виде, отвечает требованиям формализации лингвистических явлений для их последующей алгоритмической обработки.

Что касается семантической структуры исследуемых заголовочных конструкций, то она коррелируется с их структурными особенностями. При изучении семантической структуры заголовков рефератов к патентам мы опираемся на классификацию И.Е. Фроловой, выделяющую три типа номинативных предложений: констативы, директивы и экспрессивы^I. Содержанием констативов, которые в свою очередь делятся на информативы и интродуктивы, является сообщение информации. Анализ нашего материала показал, что исследуемые заголовки можно отнести к типу констативов, так как они содержат информацию о каком-либо предмете или процессе, что может быть эксплицировано следующим образом: "Данный патент содержит информацию (сообщает) о ...". Например: *Verfahren und Vorrichtung zur Drahtbefestigung an Spulen* (85,16,47,217490) - *In der Erfindung handelt es sich um ein Verfahren und eine Vorrichtung zur Drahtbefestigung an Spulen.*

В падежной грамматике указанное семантическое отношение

^I Фролова И.Е. Прагматические и семантические особенности номинативного предложения в современном английском языке: Автореф. дис. ... канд. филол. наук. - Киев, 1988. - 17 с.

обычно обозначается термином "объект" (действия, состояния, отношения). Далее для обозначения семантической функции ядерного существительного структурной модели заголовка мы будем использовать термин "объект описания". Под ним понимается наименование устройства или метода, который заявлен как патент и описывается в реферате к патенту. Все препозитивные и постпозитивные распространители ядерного существительного базовой модели заголовка рассматриваются как различные типы предикатов, иначе говоря, они представляют собой различные характеристики объекта описания.

Из изложенного следует, что в целом семантическая структура исследуемых заголовков включает два основных компонента: объект описания (способ, метод, устройство, процесс) и предикат(ы), которые внутренне дифференцируются. Они именуют: область применения устройства/метода, назначение (функцию) устройства/метода, качество (свойство) устройства/метода, особенность конструкции устройства, особенность функционирования устройства/метода, способ изготовления устройства, характеристику объекта описания с точки зрения используемого материала. Каждый предикат может быть выражен различными структурно-синтаксическими средствами.

В целях установления роли заголовка в коммуникативно-содержательной структуре текста реферата анализу подвергаются характер и способы повторов лексики заголовков в рефератах к патентам. В ходе анализа зарегистрированы следующие типы повторных номинаций: 1) полный лексический повтор; 2) повторение морфологических форм слов; 3) повторение номинации путем использования однокоренных слов; 4) повторение номинации путем перифраза и контекстуальных синонимов; 5) повторение номинации путем соотношения "часть-целое" и "целое-часть"; 6) местоименные замены; 7) вводящий повтор (полный повтор заголовка в первом предложении).

нии текста реферата к патенту). По существу данный этап анализа направлен на изучение кореференции.

Рассмотрение кореферентных связей лексики заголовка, представляющей основные компоненты его семантической структуры (объект описания и предикаты), и текста реферата, показал, что в реферате повторяются оба компонента заголовка. Однако, различные типы повторных номинаций объекта описания встречаются гораздо чаще (73%), чем заголовочных предикатов (27%), что может быть проиллюстрировано на конкретном примере (см. рис.1). Учет количественных данных повторных номинаций и анализ участия повторяемых элементов в развертывании содержания текста реферата к патенту позволяют сделать один существенный вывод: объект описания повторяется на протяжении всего текста реферата к патенту, являясь его темой. Благодаря участию разных видов повтора этого компонента семантической структуры заголовка создается так называемое тематическое поле реферата к патенту. Используя иную терминологию, правомерно утверждать, что заголовок и тематическое поле создают так называемую топикальную рамку текста реферата к патенту. В противоположность этому предикаты, которые выносятся в заголовок, выражают, по замыслу создателя текста, наиболее важные для реципиента характеристики объекта описания, вводя новую информацию о нем и тем самым образуя рематическую часть текста реферата к патенту.

Соединение номинативной и коммуникативной функций в заголовке реферата к патенту обусловило наложение его логико-семантической и коммуникативной структур. В большинстве анализируемых заголовков объект описания — название патентуемого устройства, процесса или метода — является темой озаглавливаемого реферата к патенту. А предикаты входят в рематическую часть реферата, которая охватывает практически весь текст, содержа-

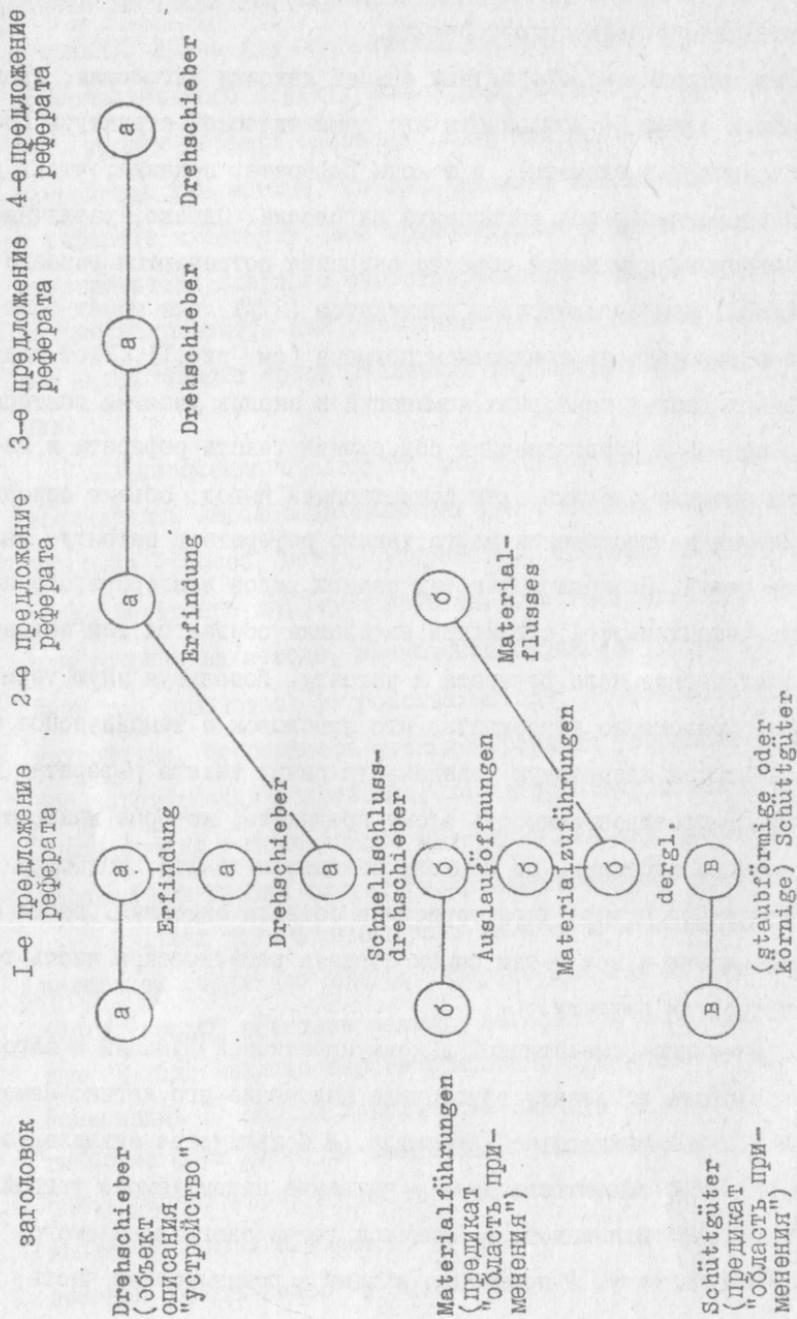


Рис.1. Кореферентные связи компонентов семантической структуры заголовка "Drehschieber in Materialführungen für Schüttgüter" и текста реферата к патенту

щий описание сути изобретения.

В диссертации проводится мысль о том, что связь заголовка с текстом реферата носит формальный и логический характер. На формальный характер связи указывает лексико-грамматическое сходство. Повторяемость заглавных слов распространяется на весь текст реферата к патенту, приобретая функцию текстовой скрепы и выстраивая формальную (внешнюю) когезию. Логическая когезия обусловлена тем, что заголовок реферата к патенту заявляет о новом изобретении, вводит информацию о нем, тогда как текст реферата ее детализирует. Повествование разворачивается от общей информации к более детальной.

Итак, заголовок реферата к патенту, являясь коммуникативной единицей, посредством которой осуществляется обмен информацией и которая содержит относительно законченное сообщение об изобретенном методе или устройстве, оказывается одновременно объектом теории коммуникации и объектом семантического синтаксиса. В свете этого план содержания заголовка предстает как состоящий из пропозиции и так называемой прагматической рамки, причем первая заключает в себе пропозициональные установки, а вторая охватывает реализационные установки на коммуникацию и прагматику.

Заголовок реферата к патенту рассматривается нами в конкретном акте коммуникации, в ходе которого он и обретает актуальность, проступающую в факторе актуального членения. Мы придерживаемся точки зрения тех лингвистов, которые считают тематическое членение одним из важных компонентов смысловой структуры предложения, а в данном случае — заголовка со статусом предложения. Именно это членение отражает не только коммуникативную природу предложения, но и познавательную установку мыслящего и говорящего субъекта. Соответственно этому принципу и были определены основные компоненты актуального членения за-

головка: исходная информация, т.е. то, что является известным или может быть легко понято, и новое, то, что сообщается автором. Исходная информация совпадает с темой озаглавливаемого текста, новое передается предикативными компонентами, образующими в совокупности рематическую часть реферата к патенту. Заголовок и передает новую информацию с опорой на известную.

По результатам анализа семантико-синтаксической структуры заголовков рефератов к патентам и связи ее компонентов с текстом реферата были выявлены следующие коммуникативные типы заголовков: 1) тема-рематический, 2) рема-тематический, 3) рема-тема-рематический, 4) тематический. Статистический анализ употребительности различных коммуникативных типов заголовков показал, что наиболее употребительным коммуникативным типом являются тема-рематические заголовки (49,8%) (см. табл.2).

Таблица 2

Употребительность коммуникативных типов заголовков рефератов к патентам по складскому хозяйству

№ п/п	Коммуникативный тип заголовка	К-во заголовков, построенных по данному типу	К-во заголовков в % к общему числу заголовков в исследуемом массиве
1.	Тема-рематические заголовки	2788	49,8
2.	Рема-тематические заголовки	1816	32,4
3.	Рема-тема-рематические заголовки	176	3,2
4.	Тематические заголовки	820	14,6
Σ	Итого:	5600	100

С учетом названных коммуникативных типов заголовков и способов их структурно-синтаксического выражения разработан алгоритм тема-рематического членения заголовка, позволяющий выделить тематический (объект описания) и рематический (предикаты) компоненты заголовка, а следовательно, и всего текста реферата.

Этот алгоритм может быть использован в информационно-поисковых системах научно-технических документов, а также в системах автоматического понимания научно-технического текста. Последнее обусловливается свойством заголовка служить семантическим центром реферата к патенту, т.е. указывать на его основные содержательные аспекты, благодаря чему и обеспечивается возможность оперативного получения сигнальной информации о патентуемых изобретениях. Таким образом, с помощью заголовков в системе МП может быть реализована важнейшая составная часть всей поисковой программы, направленной на узнавание содержательной сути патентной информации в научно-технической сфере.

Результаты исследования немецких заголовков рефератов к патентам по складскому хозяйству в плане выявления их семантико-синтаксической структуры, связи с текстом реферата к патенту, определения коммуникативных типов заголовков важны для создания системы МП немецких заголовков рефератов к патентам. Для разработанной системы МП опорой служит лингвистическая информационная база, ядренный блок которой образует автоматический словарь (АС) словоформ. Выбор способа представления лексических единиц лингвистической информационной базы в виде АС словоформ обусловлен прежде всего экспериментальным характером разработанной системы немецко-русского МП заголовков, в которой объем памяти для размещения словаря не играет решающей роли, а на первое место выдвигается быстродействие системы.

Словарная статья входного (немецкого) и выходного (русского) словников в разработанной системе МП содержит информацию, сгруппированную в четыре зоны: 1) зона машинных частей речи; 2) морфологическая зона; 3) лексико-семантическая зона; 4) зона переводных соответствий. Организация словарной статьи допускает пополнение информации как в пределах отдельных зон, так и путем

добавления новых зон.

Принцип разграничения базовых синтаксических моделей немецких заголовков на основе формальных разделителей, а также наличие необходимой информации в АС обеспечивают выделение базовых моделей алгоритмическим путем, что и определяет их использование в разработанной системе МП в качестве основных единиц машинного анализа.

Сравнительный анализ немецких заголовков и их русских переводов выявляет отсутствие однозначного соответствия выходных структур для большинства базовых моделей. При этом наибольшей неоднозначностью характеризуется двухкомпонентные базовые модели (см. табл.3).

В целях разрешения структурной многозначности при переводе с немецкого языка на русский выявляются списки диагностирующих признаков, которыми служат конкретные словоформы в определенных позициях базовых синтаксических моделей немецких заголовков.

На основе разработанного немецко-русского АС словоформ, типологии базовых синтаксических моделей немецких заголовков, их русских соответствий и списков диагностирующих признаков разработан общий и детальный алгоритмы немецко-русского МП заголовков рефератов к патентам по складскому хозяйству. Алгоритм состоит из трех основных модулей: 1) выделение очередной базовой синтаксической модели немецкого заголовка; 2) определение структуры выходной фразы на основе диагностирующих признаков; 3) выбор русских словоформ из АС с учетом специфики базовых моделей и межмодельных связей.

Программа немецко-русского перевода заголовков рефератов к патентам реализована на персональной ЭВМ и может быть использована для оперативного получения сигнальной информации о зарубежных изобретениях по складскому хозяйству. На стр.22 приведен

Структурные соответствия немецких двухкомпонентных базовых моделей в русском языке

№ п/п	Немецкие базовые синтаксические модели и их примеры	Структурные типы русских переводов и их примеры
I.	<p>S_1+S_2</p> <p><u>Abstützung eines Verbindungs-förderers^I</u> (87,14,50,241058)</p> <p><u>Verfahren und Vorrichtung zum Steuern eines Antriebssystems in einer Verpackungsmaschine</u> (87,15,49,3615442)</p> <p><u>Taben- und Flaschenverschluss-kopf mit integrierter Zeitab-Laufangabe</u> (86,4,47,3403362)</p>	<p>I. S_1+S_2 (род.п.) <u>Опора соединительного транспорта</u></p> <p>2. S_1+S_2 (твор.п.) Способ и устройство для управления системой при- вода в упаковочной маши- не</p> <p>3. $S_2 + (und) + S_1$ (род.п.) <u>Головка для закупорива- ния бутылок и трубок с встроенным индикатором времени</u></p>
2.	<p>A+S</p> <p><u>Automatische Verpackungsvorrich-tung</u> (85,15,47,3426394)</p> <p><u>Vorrichtung zum reihenweisen Überführen von flaschen</u> <u>Dauerbackwaren</u> (86,9,47,228229)</p> <p><u>Diebstahlsichere Packung</u> (86,1,47,3545775)</p> <p><u>Beladevorrichtung für kastenför-mige Grossbehälter</u> (85,15,47,3325404)</p>	<p>I. A+S <u>Автоматическое упаковоч- ное устройство</u></p> <p>2. S+A <u>Устройство для транс- портровки расположен- ных рядами плоских дол- гохранящихся кондитер- ских изделий</u></p> <p>3. S + (,) + A <u>Упаковка, препятству- ющая открытию посторо- нными лицами</u></p> <p>4. S + Adv <u>Устройство загрузки для большегрузных контейне- ров крестообразной формы</u></p>

^I В приводимых в таблице примерах подчеркнуты немецкие базовые модели и их русские переводы для каждого из выделенных типов моделей.

```

*****
ABSTREICHVORRICHTUNG FUER FOERDERBAENDER
ABSTREICHVORRICHTUNG
РАЗРАВНИВАЮЩЕЕ УСТРОЙСТВО
FUER
ДЛЯ
FOERDERBAENDER
ЛЕНТОЧНЫХ КОНВЕЙЕРОВ
РАЗРАВНИВАЮЩЕЕ УСТРОЙСТВО ДЛЯ ЛЕНТОЧНЫХ КОНВЕЙЕРОВ
*****
PALETTENBEHAELTER AUS KUNSTSTOFF
PALETTENBEHAELTER
ЕМКОСТЬ С ПОДЛОНОМ
AUS
ИЗ
KUNSTSTOFF
СИНТЕТИЧЕСКОГО МАТЕРИАЛА
ЕМКОСТЬ С ПОДЛОНОМ ИЗ СИНТЕТИЧЕСКОГО МАТЕРИАЛА
*****
DRUCKSCHLEUSE MIT KRAFTWAAGE
DRUCKSCHLEUSE
ШЛЮЗОВАЯ КАМЕРА
MIT
С
KRAFTWAAGE
СИЛОВЫМИ ВЕСАМИ
ШЛЮЗОВАЯ КАМЕРА С СИЛОВЫМИ ВЕСАМИ
*****
MASCHINE ZUR HERSTELLUNG VON VERPACKUNGSBEHAELTER
MASCHINE
МАШИНА
ZUR
ДЛЯ
HERSTELLUNG
ИЗГОТОВЛЕНИЯ
VON

VERPACKUNGSBEHAELTER
УПАКОВОЧНОЙ ТАРА
МАШИНА ДЛЯ ИЗГОТОВЛЕНИЯ УПАКОВОЧНОЙ ТАРА
*****

```

Рис.2. Фрагмент результатов машинного перевода немецких заголовков рефератов к патентам на русский язык

фрагмент результатов работы программы.

Результаты, полученные в диссертации, могут способствовать развитию дальнейших исследований, связанных с изучением семантико-синтаксической и коммуникативной структур текста и заголовка как его важного компонента. Предложенная в работе методика анализа соотношения заголовка и текста реферата к патенту на основе коррелятивных связей может найти применение при изучении взаимосвязи заголовков и текстов других жанров (научно-технических статей, описаний, монографий), что в свою очередь может быть использовано для создания алгоритмов анализа содержательной и коммуникативной структур научно-технического текста.

Основные положения диссертации отражены в следующих публикациях:

1. Автоматизированный банк данных электронного устройства-переводчика // Функционирование и развитие языковых систем. Тезисы докладов научной конференции молодых ученых, аспирантов и соискателей по итогам научно-исследовательской работы за 1976-1980 гг. - Минск, 1981. - С.159-160. В соавторстве с Т.В.Рыжаковой.
2. Пословный и поборотный анализ входных фраз в системе электронного микрокомпьютера-переводчика // Переработка текста методами инженерной лингвистики. Тезисы докладов. - Минск, 1982. - С.117-119. В соавторстве с Н.А.Дюрич, Т.В.Рыжаковой.
3. Структурная типология заголовков английских патентов как основа алгоритма машинного перевода // Системные характеристики устной и письменной речи. Сб. статей. - Минск, 1985. - С.128-134. В соавторстве с А.К.Бобковым, И.М.Мировым, Т.П.Карпилович, Л.В.Шимчук.
4. Обучение переводу заголовков научно-технических тек-

стов с помощью ЭМ // Проблемы компьютеризации процесса преподавания иностранных языков в педагогических институтах. Тезисы докладов межвузовской научной конференции. - Ульяновск, 1986. - С.56. В соавторстве с Т.П.Карпилович, И.М.Мировым.

5. О результатах опытной эксплуатации системы машинного перевода заголовков патентов // Проблемы автоматического и экспериментально-фонетического анализа текстов. Сб. статей. - Минск, 1986. - С.157-162. В соавторстве с Т.П.Карпилович, Е.В.Грановой, И.М.Мировым.

6. Моделирование переводческой деятельности в системе машинного перевода заголовков патентов // Всесоюзная школа-семинар "Бионика интеллекта". Программа и аннотации докладов. - Харьков, 1987. - С.20. В соавторстве с Т.П.Карпилович, Е.В.Грановой.

7. Компьютерное обучение грамматике немецкого языка // Лексикографические методы в обучении и научных исследованиях. Тезисы докладов областной конференции преподавателей кафедр иностранных языков. - Иваново, 1988. - С.69-70.

8. О двух системах машинного перевода заголовков патентов // Международный семинар по машинному переводу. Тезисы докладов. Тбилиси, 1989. - М.: ВЦП. - С.128-130. В соавторстве с Т.П.Карпилович.

9. Немецко-русский перевод заголовков патентов на базе ПЭМ // Актуальные проблемы компьютерной лингвистики. Тезисы докладов Всесоюзной конференции. - Тарту, 1990. - С.48. В соавторстве с Т.П.Карпилович.

10. Информативно-коммуникативная значимость научных заголовков // Использование ЭМ в научной и учебной работе гуманитарных вузов. Тезисы докладов. - Минск, 1990. - С.106-107.

8

6275

Подписано в печать.05.09.91. Формат 60x84/16. Бумага
офсетная. Объем 1,16 печ.л. Тираж 100 экз. Зак. № 86
Отпечатано на ротационте Института геохимии и геофизики
АН БССР. 220600. Минск, ул. Еодинская, 7.