81.2Ан
К 918

МИНСКИЙ ГОСУДАРСТВЕННЫЙ ЛИНГВИСТИЧЕСКИЙ УНИВЕРСИТЕТ

УДК 802.0-56

КУНЦЕВИЧ
СВЕТЛАНА ЕВГЕНЬЕВНА

СЕМАНТИКО-СИНТАКСИЧЕСКАЯ ОРГАНИЗАЦИЯ ТЕКСТОВ ТЕХНИЧЕСКИХ ОПИСАНИЙ

(на материале английского подъязыка "вычислительная техника")

Специальность 10.02.19 - общее языкознание

АВТОРЕФЕРАТ

диссертации на соискание ученой степени кандидата филологических наук

Минск, 1995

РАБОТА ВЫПОЛНЕНА В МИНСКОМ ГОСУДАРСТВЕННОМ ЛИНГВИСТИЧЕСКОМ УНИВЕРСИТЕТЕ

Научный руководитель: доктор филологических наук, профессор А. В. Зубов

Официальные оппоненты: доктор филологических наук, профессор В.В.Мартынов

кандидат филологических наук, доцент И.А.Болдак

Оппонирующая организация: БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ

Защита состоится " 21 " НОЯБРЯ **1995 г. в** 14 **час.** 00 **мин. на заседании совета по защите докторских диссертаций Д 02.22.01 при Минском государственном лингвистическом университете по адресу: 220662, г.Минск, ул.Захарова, 21.**

С диссертацией можно ознакомиться в библиотеке Минского государственного лингвистического университета.

Автореферат разослан " 17 " ОКТЯБРЯ **1995 г.**

Ученый секретарь совета по защите диссертаций, доктор филологических наук, профессор З.А.Харитончик

81.2Ан
К918

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Отсутствие в лингвистике четких эксплицитных правил представления содержания текста и его семантико-синтаксических составляющих является тормозом на пути создания более глубоких теорий жанров, подъязыков, функциональных стилей. Эта же причина не позволяет сформулировать объективные критерии для типологизации текстов. Знание эксплицитных формул текстов, основанных на их семантико-синтаксических признаках, может стать мощным инструментом для совершенствования систем автоматического анализа текстов (индексирования, реферирования, аннотирования, машинного перевода) и создания компьютерных систем порождения научно-технических документов различных жанров. Все сказанное в достаточной степени определяет **актуальность** выполненного исследования, которое, главным образом, посвящено построению и изучению семантико-синтаксических формул отобранных для анализа текстов.

Как известно, научно-технические тексты несут основную информацию, влияющую на прогресс общества в целом. Однако в лингвистике практически отсутствуют крупные работы, анализирующие структуру и содержание связных текстов какого-либо научно-технического подъязыка с точки зрения их представления в компьютере. Поэтому **актуальным** является и "машинноориентированный" подход к формальному представлению содержания таких текстов.

Основным **предметом** нашего **исследования** выступают формальные правила семантико-синтаксической организации английских текстов технических описаний.

Материалом исследования послужил массив из 51 текста, длиной от трех до восемнадцати абзацев каждый. Указанные тексты, отнесенные нами к жанру *техническое описание*, были отобраны из англоязычной американской и канадской специальной литературы по вычислительной технике (научно-технические журналы "Datamation", "Mini-Micro Systems", "Computer" и "PC Today" за 1990г.; учебное пособие по информатике для студентов и преподавателей университетов США -- Larry Long and Nancy Long "Computers" 1986 года издания). В исследованном корпусе текстов выявлены две жанровые разновидности технических описаний. С учетом коммуникативного задания этих групп текстов их можно определить как *рекламно-технические описания* (РТО) и *научно-популярные технические описания* (НПТО).

Целью предлагаемого исследования является проверка гипотезы относительно существования для текстов достаточно узкой предметной области, относящихся к одному и тому же жанру научно-технической литературы, некоторых семантико-синтаксических моделей, записанных на специальном языке, понятном компьютеру, и построение эксплицитных формул этих текстов.

В процессе достижения поставленной цели решались следующие **основные задачи:**

1. Определение релевантных в плане настоящего исследования свойств текста как самостоятельно функционирующей единицы коммуникации.

БІБЛІЯТЭКА

2. Выявление основных единиц научно-технического текста на тематическом, логико-семантическом и семантико-синтаксическом уровнях и анализ их межуровнего соответствия.
3. Исследование тематической структуры текстов технических описаний подъязыка "вычислительная техника" (ВТ).
4. Описание исследуемых текстов на логико-семантическом уровне.
5. Создание семантико-синтаксического языка для формального представления содержания научно-технических текстов.
6. Разработка методики описания семантико-синтаксической организации значимых составляющих отдельных текстов и текстов в целом.
7. Исследование семантико-синтаксической структуры двух жанровых разновидностей текстов технических описаний подъязыка ВТ и выявление наиболее типичных семантико-синтаксических формул указанных текстов.

Используемые **методы исследования:** семантико-синтаксическое, тематическое, ситуативное и логико-семантическое моделирование; сопоставительный, контекстологический, трансформационный и статистический анализ.

В соответствии с целями и задачами исследования на защиту выносятся следующие **основные положения**:

1. Основным содержательным компонентом текстов, принадлежащих к жанру *техническое описание*, является абзац.
2. С опорой на теорию семантических падежей (функций) может быть создан семантико-синтаксический язык СЕМСИНТ/А, понятный компьютеру и позволяющий ему построить семантико-синтаксическую формулу текста (СЕСФТ) в виде цепочки эксплицитных семантико-синтаксических формул абзацев (СЕСФА).
3. Минимальной составляющей содержательной структуры абзаца является элементарный компонент содержания (ЭКС).
4. В текстах одного жанра в рамках конкретного научно-технического подъязыка существует конечное число ЭКС.
5. В текстах одного жанра в рамках конкретного научно-технического подъязыка можно выделить ограниченное число канонических моделей (КМ) абзацев как предельно обобщенных СЕСФА.
6. Тексты одного жанра имеют конечное число обобщенных семантико-синтаксических моделей.
7. Обобщенные семантико-синтаксические модели текстов одного жанра строятся из КМ, специфичных для данного жанра, на основе вероятностных и детерминированных правил порождения текстов.

Научная новизна проведенного исследования заключается в том, что впервые в рамках лингвистики текста разработан комплексный подход к моделированию содержания научно-технических текстов, учитывающий композиционные, грамматические и

семантические особенности этих текстов. Названный подход является развитием и совершенствованием метода формального представления содержания текста, предложенного в исследованииях А.В.Зубова для русского языка. Создан новый семантико-синтаксический язык, позволяющий формализовать содержание англоязычных научно-технических текстов. Дальнейшая разработка впервые введеных автором диссертации понятий *ЭКС абзаца*, *КМ абзаца*, *обобщенная семантико-синтаксическая модель текста* (ОССЕМТ) могла бы внести определенный вклад в общую теорию языкознания.

Теоретическое значение настоящего исследования состоит в том, что в нем предложен механизм сопоставления семантико-синтаксической структуры значимых составляющих текста и целых текстов, а также извлечения из текстов основной и факультативной информации. Полученные результаты позволяют уточнить и конкретизировать роль абзаца в структуре научно-технического текста. Выявленные в ходе исследования формулы абзацев двух жанровых разновидностей текстов технических описаний английского подъязыка ВТ дают новые эксплицитные критерии различения речевых жанров и тем самым способствуют уточнению некоторых основных понятий функциональной стилистики. Полученные семантико-синтаксические формулы абзацев и текстов могут служить основой для создания новых приемов типологических исследований, базирующихся на более объективных, чем существующие ныне, критериях.

Практическое значение результатов исследования. Предложенная автором методика формального представления содержания научно-технических текстов может быть с успехом использована и при анализе иных типов английских текстов. Представленные в диссертации семантико-синтаксические формулы текстов -- это готовый материал для создания компьютерных программ порождения англоязычных текстов технических описаний двух исследованных жанровых разновидностей. Разработанные на аналогичном более широком материале компьютерные программы подобного рода могут стать коммерческим продуктом, способным автоматически создавать технические описания на готовую промышленную продукцию (компьютеры, телевизоры, телефоны и т.п.).

Апробация работы. Материалы работы апробированы в докладах на заседаниях кафедры информатики и прикладной лингвистики Минского государственного лингвистического университета (МГЛУ), на научной конференции этого университета по итогам научно-исследовательской работы за 1989г., на Всесоюзной конференции "Использование ЭВМ в научной и учебной работе гуманитарного вуза" (Минск, 1990г.), на республиканских конференциях "Интенсивные методы обучения иностранным языкам" (Минск, 1991г.), "Философия. Язык. Культура" (1991г.), "Беларуская мова сярод еўрапейскіх" (1993г.), "Компьютерные программы в обучении белорусскому и иностранным языкам" (1994г.).

Основные положения данного исследования отражены в шести публикациях.

Структура настоящей **работы** определяется целями и задачами проведенного исследования. Диссертация состоит из введения, трех глав, заключения, библиографии, списка использованных словарей и приложений.

Во **Введении** обосновывается актуальность темы исследования, определяется ее научная новизна, теоретическая и практическая значимость, формулируются основные цели и задачи исследования, а также указываются основные методы анализа, которые использовались в процессе обработки лингвистического материала.

Первая глава содержит теоретический анализ работ по проблемам связного текста, рассматривающих его основные характеристики в свете коммуникативно-функционального подхода к языку. Здесь же определяется роль абзаца в тематической, логико-семантической и композиционной структуре текста.

Во **второй главе** проводится анализ тематической и логико-семантической структуры отобранных текстов. В этой части работы описывается семантико-синтаксический язык СЕМСИНТ/А, созданный нами для формального представления содержания исследуемых текстов, а также приводятся правила описания этих текстов средствами упомянутого языка на уровне абзацев.

В **третьей главе** исследуются основные типологические черты отобранных для анализа разновидностей текстов технических описаний, изучается специфика текстов РТО и НПТО в рамках английского подъязыка ВТ, а также приводятся наиболее типичные обобщенные семантико-синтаксические модели указанных текстов и описывается методика их выявления.

В **Заключении** обобщаются результаты исследования и намечаются перспективы его дальнейшего развития.

Полный объем настоящей диссертации составляет 146 страниц. Таблицы занимают 41 страницу основной части работы и позволяют проиллюстрировать и более полно отразить разнообразные наблюдения, полученные в процессе обработки лингвистического материала. В трех приложениях приведены тексты и абзацы, ссылки на которые содержатся в диссертации, а также некоторые сопутствующие результаты анализа исследованного корпуса текстов.

СОДЕРЖАНИЕ РАБОТЫ

Исследование связного текста как особого речевого образования с единым, самостоятельным содержанием является одной из актуальных задач современного языкознания.

Появление лингвистики текста подготовлено и обусловлено идеей о необходимости различения языка и речи, известной в языкознании со времен В. Гумбольдта и четко сформулированной Ф. де Соссюром. При изучении языка как коммуникативной системы

неизбежен выход в область его реального функционирования в различных сферах и ситуациях общения. Именно в речи через последовательные серии связанных между собой предложений, которые и образуют в конечном счете текст, непосредственно отражается движение мысли говорящего или пишущего. Естественный язык функционирует только в текстах (больших и малых, устных и письменных), посредством которых осуществляется само общение между людьми.

Как уже показано в многочисленных исследованиях, для текста как самостоятельной коммуникативной единицы характерны две следующие особенности:

I. Текст является целостным и связным лингвистическим образованием как с точки зрения своего содержания, так и чисто формально.

II. Правила порождения текстов и их единиц (абзацев, предложений,словосочетаний) подчиняются законам функционирующей языковой системы.

Раскрывая эти особенности подробнее, отметим релевантные для нашего исследования характеристики научно-технического текста:

1) графическая объективированность;

2) наличие единого содержания;

2) целостное описание некоторого фрагмента действительности;

3) наличие различных структурных единиц, в качестве которых выступают: слово, словосочетание, предложение и абзац;

4) наличие логической, грамматической и семантической связи между единицами текста;

5) наличие модальности и прагматической установки;

7) наличие основной семантико-синтаксической единицы, в качестве которой выступает абзац.

В реферируемой диссертации разработан комплексный подход к анализу содержания научно-технических текстов, включающий три основных уровня представления их содержательной структуры: логико-семантический, тематический и семантико-синтаксический. Изучены некоторые межуровневые соответствия.

В основу проведенного нами исследования **логико-семантической организации** текстов положено понятие *функционально-смысловой тип абзаца* (ФСТА), поскольку логеме мыслительного процесса крупнее элементарного суждения, обычно соотносимого с предложением, на языковом уровне мы ставим в соответствие именно **абзац** научно-технического текста.

Анализ соответствующей литературы позволяет определить **ФСТА** как образец, модель фрагмента связного текста, с определенным набором языковых средств выражения его обобщенного значения. Как правило, выделяются основные и смешанные ФСТА. К первым относятся *абзац-описание* (О), *абзац-рассуждение* (Р) и *абзац-повествование* (П). Смешанные ФСТА формируются на базе основных ФСТА (например,

описание/рассуждение -- О/Р, *описание/повествование* -- О/П и т.п.) , представляя собой при этом некоторую самостоятельную данность.

В текстах РТО нами выявлены следующие ФСТА: *описание*, *описание/рассуждение* и *описание/повествование*. В начальных абзацах текстов указанной жанровой разновидности абзацы смешанных типов не зафиксированы. При этом абзац *О/Р* более характерен для медиальной части этих текстов, а абзац *О/П* -- для их финальных абзацев.

Анализ логико-семантической структуры текстов РТО свидетельствует об их явной "описательности", поскольку 85% всех абзацев этих текстов представляют собой *описания*, а оставшиеся 15% абзацев смешанного типа можно охарактеризовать как *абзацы-описания с элементами рассуждения или повествования*.

Линейную последовательность ФСТА каждого текста можно представить в виде формулы, которую далее будем называть **логико-семантической формулой текста**. Так, логико-семантическая формула значительной части обследованных текстов РТО имеет вид:[1]

$$Tij=O_н \,\&\, O_2 \,\&\ldots\&\, O_n \,\&\, O_ф$$

В результате анализа выяснилось, что логико-семантическая структура текстов НПТО более разнообразна. Наряду с ФСТА, выделенными в текстах РТО (О, О/Р и О/П), в текстах НПТО обнаружены: *абзац-рассуждение*, *абзац-повествование* и абзац типа *повествование/рассуждение* (П/Р), причем, число последних в общем корпусе текстов НПТО незначительно.

Однако, и в текстах НПТО абзацы описательного характера (О, О/Р, О/П) преобладают, составляя в сумме 91% от общего количества абзацев, хотя предпочтение отдается не "чистому" *описанию* -- *О*, как в текстах РТО, а *описанию с элементами рассуждения* -- *О/Р*.

Таким образом, тексты НПТО также можно отнести к жанру *техническое описание*. Причем специфику организации текстов этой жанровой разновидности (в сравнении с текстами РТО, в них присутствует более широкий набор ФСТА и преобладают абзацы типа *О/Р*) определяет, на наш взгляд, именно научно-популярный способ изложения материала.

Характеризуя композиционную организацию текстов НПТО, укажем, что их начальные абзацы по своей функционально-смысловой принадлежности относятся, в основном, к типу *О/Р*, медиальные -- к абзацам *О/Р* и *О*; в финальной части этих текстов к названным ФСТА прибавляются абзацы типа *О/П*. Исходя из сказанного, обобщенную логико-семантическую формулу текстов технических описаний научно-популярного характера можно

---

[1] В логико-семантических формулах текстов, Tij соответствует номеру текста в общем массиве технических описаний; индекс "н" обозначает начальный абзац текста, "ф" -- финальный; цифровые индексы нумеруют абзацы в порядке их следования друг за другом в тексте. Знак "&" отмечает связь между абзацами текста; знак "V" обозначает необходимость выбора одного из указанных ФСТА: О -- абзац типа *описание*, О/Р -- *описание/рассуждение*, О/П -- *описание/повествование*.

представить следующим образом:

$$Tij=O/P_н \,\&\, C/P_2 \,\&/\vee\, O_3 \,\&/\vee\, \ldots\, O/P_n \,\&\, (O_ф \vee O/P_ф \vee O/П_ф)$$

Однако, приведенные здесь формулы отражают лишь общее логико-семантическое содержание проанализированных текстов. Чтобы формализовать полное содержание этих текстов, необходимо исследовать их тематическую структуру и разработать специальный язык, который бы позволил детально описать семантику и синтаксис отдельных абзацев и текстов в целом.

Базовым понятием двух других уровней представления наших текстов (**тематического и семантико-синтаксического**) выступает *ЭКС абзаца*.

ЭКС абзаца научно-технического текста назовем минимальную инвариантную составляющую тематической структуры абзаца, которая в тексте, записанном на естественном языке, может репрезентироваться частью предложения, целым предложением, несколькими предложениями абзаца или даже полным абзацем текста. Один и тот же ЭКС выражается сходными семантико-синтаксическими структурами и набор подобных структур для каждого ЭКС конечен.

Выделение этой новой для лингвистики текста единицы осуществляется методом бинарных оппозиций абзацев, обнаруживающих определенное сходство своего тематического содержания, по аналогии с тем, как в структуре значения слова выделяются семы. Путем сопоставления содержания анализируемых абзацев друг с другом удалось установить список тематических компонентов, формирующих структуру их содержания. Это же позволило сгруппировать абзацы на основании наличия/отсутствия в них того или иного компонента тематического содержания.

Различаются ЭКС объективного характера (ЭКС I порядка) и ЭКС субъективного характера (ЭКС II порядка). Первые связаны с представлением основного содержания научно-технического текста, вторые -- с его факультативным содержанием.

Основное содержание текста определяется совокупностью линейных цепочек ЭКС I порядка, зафиксированных для каждого из его абзацев. ЭКС II порядка, описывающие факультативное содержание текста, учитываются только в формуле полного тематического содержания текста.

Была исследована гипотеза о том, что тексты достаточно узкой предметной области и одинаковой жанровой принадлежности имеют конечный список ЭКС и что конечен также набор их возможных комбинаций. В ходе исследования текстов технических описаний английского подъязыка ВТ (см.таблицу 1) было выделено двенадцать ЭКС основного содержания текста и восемь ЭКС факультативного содержания текста. Вместе с тем изучены закономерности сочетаемости ЭКС как с точки зрения аспектов содержания указанных текстов, так и с точки зрения их композиционной структуры. Полученные результаты подтверждают выдвинутую гипотезу.

В отобранном корпусе текстов выявлены 73 возможные комбинации ЭКС для абзацев текстов РТО и 103 -- для абзацев текстов НПТО. В результате анализа этих комбинаций

нами была выделена 91 линейная цепочка ЭКС I порядка, каждая из которых отражает фрагменты основного тематического содержания обеих разновидностей текстов. Указанные цепочки ЭКС рассматриваются нами как микротемы (Mnnn) текстов технических описаний. Наиболее употребительные из них включены в перечень микротем, встречающихся в обеих жанровых разновидностях текстов технических описаний подъязыка ВТ от 25 до 9 раз:

M003=ЭКС03
M023=ЭКС02&ЭКС04
M013=ЭКС08
M110=ЭКС05
M016=ЭКС09
M023=ЭКС03&ЭКС05
M048=ЭКС02&ЭКС03&ЭКС04

Для формального описания семантико-синтаксической организации исследуемых английских текстов разработан специальный семантико-синтаксической язык (СЕМСИНТ/А). В предлагаемый нами язык входят следующие составляющие:

1) алфавит языка СЕМСИНТ/А,
2) система средств для записи семантических отношений между словами предложения,
3) система средств для записи синтаксических отношений между словами предложения,
4) система элементов для фиксации логико-семантических отношений между предложениями абзаца и абзацами текста.

Наряду со средствами кодирования англоязычных текстов указанный язык включает правила и приемы построения и содержательной обработки семантико-синтаксических формул как отдельных абзацев, так и целых текстов.

Алфавит языка СЕМСИНТ/А включает в себя все знаки основного набора клавиатуры современных персональных компьютеров. Комбинации букв, цифр и специальных знаков алфавита позволяют описывать все остальные составляющие нашего языка.

В основу двух следующих составляющих языка СЕМСИНТ/А положена глубинно-падежная концепция представления смысла предложения, впервые сформулированная Ч.Филлмором. Анализ идеи Ч.Филлмора, его последователей, а также учет специфики исследуемых нами текстов дали возможность выделить 52 семантические функции (СФ), определяющие семантические отношения между частями предложения. Такими частями являются в нашей модели аргументные группы (АГ). На поверхностном уровне АГ может совпадать либо с группой слов, либо с отдельным словом предложения. В языке СЕМСИНТ/А аргументные группы заключаются в угловые скобки [< >], перед которыми ставится код соответствующей семантической функции.

Таблица1

ЭКС текстов РТО и НПТО[1]

N п/п	Наименование элементарных компонентов содержания (ЭКС) абзацев	Тип текстов		Код ЭКС
		РТО	НПТО	
1	2	3	4	5
1	Сообщение о появлении или существовании аппаратных и/или программных средств (АПС)	+	+	**ЭКС01**
2	Состав АПС	+	+	ЭКС02
3	Функциональные возможности АПС	+	+	ЭКС03
4	Количественные технические характеристики АПС	+	+	ЭКС04
5	Назначение или область применения АПС	+	+	ЭКС05
6	Совместимость АПС с существующими системами	+	+	ЭКС06
7	Конструктивные особенности АПС	+	+	ЭКС07
8	Преимущества АПС или технологического процесса, связанного с ВТ, над существующими аналогами	+	+	ЭКС08
9	Принцип работы АПС или ход технологического процесса, связанного с ВТ	+	+	ЭКС09
10	Процедуры пользователя по эксплуатации АПС	+	+	ЭКС10
11	Маркетинговая политика фирм(ы) относительно того или иного АПС	+	+	**ЭКС12**
12	Толкование понятия ВТ	-	+	ЭКС13
13	Состояние АПС или процесса, связанного с ВТ	-	+	**ЭКС14**
14	Средство поддержки диалога с получателем текста	-	+	**ЭКС15**
15	Недостатки АПС или технологического процесса, связанного с ВТ	-	+	ЭКС16
16	Сходство АПС или технологических процессов, связанных с ВТ	-	+	ЭКС17
17	Различие АПС или технологических процессов, связанных с ВТ	-	+	ЭКС18
18	Классификация АПС	-	+	**ЭКС20**
19	Значение АПС или технологического процесса, связанного с ВТ	-	+	**ЭКС21**
20	Причины появления публикации	-	+	**ЭКС22**

[1] В таблице 1 жирным шрифтом отмечены ЭКС II порядка. Коды ЭКС11 и ЭКС19, введенные на начальном этапе анализа, в результате уточнений из общего списка исключены.

Важно отметить, что предлагаемые нами СФ неравноправны с точки зрения синтаксической структуры предложения и выступают как актанты, сирконстанты или детерминативы.

Принадлежность слова (словосочетания) к той или иной семантической группе отражена в кодах СФ **актантов** (хотя имеются и некоторые другие средства). Так, например, *активный неодушевленный инициатор события* может представлять собой либо *предмет* (AGN), либо *материал* (AGM), либо *характеристику* (AGH), либо *процесс* (AGP), либо относиться к *общим* (AGG) или *научным* (AGC) *понятиям*. Указанные СФ можно привести к единому виду, заменив кодовое обозначение семантической принадлежности слова (последние буквы кодов СФ -- N, M, H, P, G, C) на X. Полученная функция AGX называется **приведенной** и является в данном случае маркером *неодушевленного инициатора события*. Для того, чтобы определить, к какой конкретно семантической группе (N, M, H, P, G или C) относится анализируемое слово (группа слов), мы используем, по возможности, формальные диагностирующие контексты.

К **сирконстантным СФ** относятся, например, LOC -- *место совершения некоторого события*, FRQ -- *частота совершения некоторого события*, FRW -- *исходная точка движения*, TOW -- *конечная точка движения* и т.п.

Детерминативами в нашем языке выступают СФ PRP -- *свойство*, SPC -- *уточнение*, QNT -- *количество*, DSN -- *назначение* и некоторые другие.

Если семантическая функция выражена синтаксически отдельным предложением, то это помечается штрихом [']. Например: TIM' маркирует придаточное предложение времени, ONX' -- дополнительное придаточное предложение, PRP' -- определительное придаточное предложение и т.д.

Для обозначения в семантико-синтаксической структуре предложения места глагола-сказуемого используется код Rn, где n -- цифра, отражающая семантическую валентность глагола. Семантическая валентность глагола определяется количеством аргументных групп, которые данный глагол объединяет в каждом конкретном случае. Причем, маркер R в семантико-синтаксической формуле предложения далеко не всегда совпадает с грамматическим сказуемым в традиционном понимании.

Все составные части семантико-синтаксической формулы предложения (АГ и глагол-сказуемое) соединяются между собой знаком "плюс" [+] в строгом соответствии с порядком следования слов в предложении и с сохранением знаков пунктуации текста.

Любое предложение или цепочку предложений на языке СЕМСИНТ/А можно записать несколькими способами. Например, английское предложение *Stardent's 3000 machine doubles the processing power of its predecessor* при детальном представлении его семантико-синтаксической структуры будет выглядеть так:

AGN<N013*NUM1*N074>+R2<V009>+ONH<art3*Vng<V006>*N072*d090*p017*N074>.

В предлагаемой формуле N013, N074, N072; NUM1, V009, art3, d090, p017 -- коды семантических подклассов существительных; числительных, глаголов, артиклей, предлогов

и местоимений соответственно. Запись "Vng<V006>" обозначает *ing*-форму глагола, принадлежащего к семантическому подклассу V006. Знак "умножения" [*] связывает в единое целое компоненты отдельной АГ.

Другой способ позволяет записать взятое для примера предложение в более обобщенном виде:

AGN+R2<V009>+ONH.

Эта запись занимает мало места и ее легко прочесть (*Предмет улучшает некоторую характеристику.*), если иметь в виду, что функция AGN обозначает активного неодушевленного инициатора события типа *предмет*, ONH есть объект события типа *характеристика*, а глагол-сказуемое с кодом V009 используется для выражения преимуществ (группа глаголов со значением 'улучшать', 'убыстрять', 'ускорять' и т.п.) аппаратных и/или программных средств ВТ над существующими аналогами. Названное обстоятельство делает этот вариант семантико-синтаксического кодирования более приемлемым для настоящего исследования, не ставящего целью порождение научно-технических текстов.

Английские научно-технические тексты изобилуют неличными формами глаголов и предикативными конструкциями с ними. В предлагаемом языке мы выделяем инфинитив (Vnf), причастие прошедшего времени (Ved) и единую *ing*-форму глагола (Vng). В языке СЕМСИНТ/А предусмотрены средства кодирования названных глагольных форм с учетом их предикативной потенции. Кроме них, как показал анализ, предикатами второго порядка могут быть также процессные существительные (Rpr) и прилагательные (Raj), выступающие в роли именной части традиционного сказуемого. Например:

а) *A UNIX spreadsheet program provides support for the X-Windows system windowing environment.*⇒ AGN+R2+ONP<**Rpr1**+DDM>.

б) *Exclaim will be available at the end of the second quarter of 1990.*
⇒ ANG+R2+PRP<**Raj1**+TIM>.

Специалистами по лингвистике текста проанализированы и классифицированы различные логико-семантические средства связи, существующие между предложениями и абзацами письменного текста (**четвертая составляющая** языка СЕМСИНТ/А). Их можно представить как логико-семантические скрепы (в скобках записаны коды, закрепленные за ними в нашем языке), выражающие: субъективное отношение автора к сообщаемому в тексте (lk01); причинно-следственные связи (lk02); перечисление (lk03); возражение (lk04); противопоставление (lk05); подтверждение (lk06); результат (lk07); итог (lk08); обстоятельство (lk09); оценку достоверности сообщения (lk10) и т.п.

Описанные выше составляющие языка СЕМСИНТ/А позволяют формально представить семантико-синтаксическую структуру как основных единиц текста, так и текста в целом.

Как было отмечено выше, основным семантико-синтаксическим компонентом научно-технического текста является абзац. Его содержание на языке СЕМСИНТ/А записывается в

виде *семантико-синтаксической формулы абзаца* (СЕСФА), т.е. СЕСФА есть формализованное представление содержания абзаца. Мы различаем четыре основных типа СЕСФА, которые перечислены здесь по степени возрастания уровня их абстрагирования: *рабочая СЕСФА*, *полная СЕСФА*, *обобщенная СЕСФА* и *каноническая СЕСФА*.

Наиболее детально семантико-синтаксическую структуру абзаца описывает полная СЕСФА. Рабочая СЕСФА несколько приближена к реальному тексту, поскольку предполагает запись служебных слов на естественном языке. Обобщенная СЕСФА содержит лишь коды необходимых СФ. Она кратко описывает семантико-синтаксическую структуру абзаца и дает возможность **формально** подойти к выявлению его тематической организации. Каноническая СЕСФА является пределом абстрагирования обобщенной СЕСФА и позволяет сравнивать семантико-синтаксическое содержание абзацев на самом высоком уровне обобщения.

Минимальным же компонентом содержания **абзаца** в рамках реферируемой работы считается ЭКС. Одна из наших задач заключалась в том, чтобы показать какие элементы семантико-синтаксической структуры абзаца, представленного на языке СЕМСИНТ/А, сигнализируют о наличии в нем того или иного ЭКС. Такими элементами могут быть характерные структуры предложений или их частей, характерная семантическая функция, характерный семантический подкласс глагола и т.п.

Процедура анализа СЕСФА с одинаковым набором ЭКС заключается в их сопоставлении друг с другом с учетом следующих условий:

1. Если обобщенные СЕСФА двух или более абзацев полностью **идентичны**, то можно говорить об их семантико-синтаксическом **тождестве**.
2. Если обобщенные СЕСФА удается привести к единому **каноническому виду**, то они проявляют семантико-синтаксическое **подобие**.
3. Если обобщенные СЕСФА имеют хотя бы один **формальный признак** ЭКС на семантико-синтаксическом уровне, то налицо **толерантность** СЕСФА.
4. Если внешне **обобщенные** СЕСФА совершенно различны, то необходимо проанализировать их **полные** СЕСФА.

Обобщенную СЕСФА, над которой были выполнены хотя бы некоторые операции приведения к каноническому виду, назовем **приведенной СЕСФА**. Общий вид приведенной СЕСФА зависит от количества преобразований, которые над ней были произведены. Приведенная СЕСФА занимает как бы промежуточную ступень между обобщенной и канонической СЕСФА.

Сравнение семантико-синтаксической структуры абзацев осуществлялось нами, прежде всего, на уровне обобщенных СЕСФА. Выход же из глубинного уровня к поверхностному использовался достаточно редко. Это делалось лишь тогда, когда формальные признаки интуитивно отмеченных в абзаце ЭКС отсутствовали в обобщенной СЕСФА.

В работе сформулированы **правила записи** обобщенных СЕСФА и основные **правила приведения СЕСФА** к каноническому виду.

В настоящем исследовании **каноническая модель абзаца** с той или иной микротемой "выводится" из соответствующих обобщенных семантико-синтаксических формул абзацев исследуемых текстов. Канонические модели абзацев выступают тем строительным материалом, из которого, в соответствии с нашей гипотезой, формируются семантико-синтаксические модели текстов узкой предметной области и одинаковой жанровой принадлежности. Список таких моделей конечен, хотя бы потому, что принципиально конечен и список отражаемых в них типовых ситуаций.

Следующим этапом нашего анализа было сопоставление **семантико-синтаксических формул текстов** (СЕСФТ) двух исследуемых жанровых разновидностей текстов *технических описаний*, имеющих сходную тематическую структуру.

В процессе выполнения настоящего исследования было изучено 33 текста РТО, при сопоставлении которых **создавалась** методика описания целых текстов, и 18 текстов НПТО, на которых разработанная методика **проверялась**.

Процедура сопоставления СЕСФТ близких по содержанию текстов включает следующие этапы:

1) Запись семантико-синтаксических формул основного содержания (СЕСФОСТ) исследуемых текстов.
2) Проверка СЕСФОСТ на наличие в них канонических моделей абзацев, соответствующих искомым микротемам, или цепочкам ЭКС I порядка.
3) Приведение СЕСФОСТ к каноническим моделям входящих в их состав абзацев.
4) Сжатие СЕСФОСТ за счет удаления повторяющихся канонических СЕСФА и запись полученных в результате выполнения этой операции приведенных СЕСФОСТ.
5) Сжатие приведенных СЕСФОСТ за счет дополнительного кодирования их сходных фрагментов, которые назовем **каноническими фрагментами текстов** (КФТ), и запись обобщенной СЕСФТ (ОСЕСФТ).
6) Выявление набора ОСЕСФТ, при помощи которых можно представить значительную часть всех анализируемых текстов.

Назовем семантико-синтаксические формулы текстов, включенные в множество ОСЕСФТ, описывающих подавляющее большинство текстов, **обобщенными семантико-синтаксическими моделями текстов** исследуемой жанровой разновидности.

Анализ показывает, что обобщенными моделями текстов РТО можно признать, по крайней мере, восемь их ОСЕСФТ. Что же касается текстов НПТО, то 83% текстов этой жанровой разновидности технических описаний покрываются тремя ОСЕСФТ, а для описания всего массива указанных текстов достаточно шести ОСЕСФТ. На основании принятого нами допущения, наиболее продуктивные ОСЕСФТ будем считать обобщенными семантико-синтаксическими моделями текстов НПТО.

Сравнение ОСЕСМТ обеих разновидностей текстов между собой позволяет утверждать, что тексты РТО характеризуются большей однородностью своей семантико-синтаксической структуры на уровне входящих в их состав канонических моделей абзацев или, в более обобщенном варианте, канонических фрагментов текста, поскольку их ОСЕСМТ "компактнее", чем ОСЕСМТ текстов НПТО.

На наш взгляд, вышесказанное подтверждает выдвинутую в настоящей работе гипотезу относительно принципиальной возможности представления множества текстов одного жанра в масштабе того или иного научно-технического подъязыка при помощи сравнительно небольшого набора семантико-синтаксических моделей. Однако следует иметь в виду, что сами модели подвержены некоторым изменениям как ввиду внутреннего развития языка, так и под влиянием внеязыковых факторов.

Эксплицитными характеристиками жанра *техническое описание* на исследованном нами массиве текстов выступают 11 наиболее употребительных канонических моделей абзацев, на базе которых можно построить 11 ОСЕСМТ, описывающих содержание подавляющего большинства этих текстов.

Рассматривая выполненное нами исследование в более широком плане, необходимо выделить следующие направления его возможного развития:

1. Расширение объема анализируемых текстов за счет привлечения технических описаний других подъязыков (например, оптики, аудио- и видеотехники и т.п.). Это даст возможность детально описать специфику текстов указанного жанра на семантико-синтаксическом уровне и перейти к их компьютерному созданию.

2. Исследование текстов оригинальных статей, монографий, учебников и пособий по ВТ с целью построения семантико-синтаксической модели английского подъязыка *ВТ*.

3. Привлечение для исследования по разработанной методике текстов различных подъязыков и выявление всех разновидностей текстов жанра *техническое описание*. Их сопоставительный анализ на уровне канонических моделей абзацев и обобщенных семантико-синтаксических моделей текстов позволит внести значительный вклад в общую теорию лингвистики текста.

Основные положения диссертации отражены в следующих **публикациях автора**:

1. Зноско С.Е. К вопросу о взаимодействии лексической семантики и синтаксиса (на материале подъязыка вычислительной техники) // Депонированный сборник кафедры лексикологии английского языка МГПИИЯ "Функционально-семантические аспекты лексики в германских языках". -- Минск, 1987. -- С. 80 -- 86.

2. Удинская И.П., Карпилович Т.П., Зноско С.Е. Автоматизированный англо-русский словарь для АРМ переводчика АМПЕР // Международный семинар по машинному переводу. Тбилиси, 27 ноября -- 2 декабря, 1989 г. Тезисы докладов. -- Москва: Всесоюзный центр переводов. -- С. 307 -- 309.

3. Удинская И.П., Кунцевич С.Е. Англо-русский автоматизированный словарь по вычислительной технике // Использование ЭВМ в научной и учебной работе гуманитарного вуза. Тезисы докладов Всесоюзной конференции 12 -- 14 июня

1990 г. -- Минск, 1990. -- С. 123 -- 124.

4. Кунцевич С.Е., Удинская И.П. Возможности использования англо-русского автоматизированного словаря по вычислительной технике // Интенсивные методы обучения иностранным языкам. Материалы научно-технической конференции 25 -- 27 июня 1991 года. -- Минск: РТИ, 1991. -- С. 90.

5. Кунцевич С.Е. Комплексное моделирование содержания текстов технических описаний как основа создания обучающих программ работы с текстами // Компьютерные программы в обучении белорусскому и иностранным языкам. Республиканская конференция 14 -- 15 июня 1994 г. Тезисы докладов. -- Минск: МГЛУ, 1994. -- С. 28 -- 30.

6. Кунцевич С.Е. О тематической структуре английских текстов рекламно-технических описаний // Беларуская мова сярод еўрапейскіх. Зборнік тэзісаў. -- Мінск: МДЛУ, 1994 . -- С. 56 -- 60.

7. Кунцевич С.Е., Удинская И.П. Семантико-синтаксический язык для формального представления смысла текста // Философия. Язык. Культура. В 5 частях. Часть 3. -- Минск, 1991 г. -- В печати.

Р Е З Ю М Е

КУНЦЕВИЧ СВЕТЛАНА ЕВГЕНЬЕВНА

СЕМАНТИКО-СИНТАКСИЧЕСКАЯ ОРГАНИЗАЦИЯ ТЕКСТОВ ТЕХНИЧЕСКИХ ОПИСАНИЙ (НА МАТЕРИАЛЕ АНГЛИЙСКОГО ПОДЪЯЗЫКА "ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА")

Научно-технический текст, жанр, техническое описание, содержание текста, логико-семантическая структура текста, тематическая структура текста, семантико-синтаксическая структура текста, элементарный компонент содержания абзаца, семантико-синтаксический язык для формального представления содержания текста, основное содержание текста, факультативное содержание текста, правила преобразования семантико-синтаксических формул абзацев, каноническая модель абзаца, обобщенная семантико-синтаксическая модель текста.

Исследование выполнено с целью проверки гипотезы относительно существования для текстов достаточно узкой предметной области и одинаковой жанровой принадлежности конечного набора их обобщенных семантико-синтаксических моделей. Разработан новый комплексный подход к описанию содержания научно-технических текстов, учитывающий три уровня представления текста: логико-семантический, тематический и семантико-синтаксический. Предложен способ извлечения основной и факультативной информации из научно-технического текста. Введена новая единица анализа научно-технического текста -- элементарный компонент содержания (ЭКС) абзаца. Создан новый семантико-синтаксический язык для формальной записи содержания английского научно-технического текста. Описана методика создания и сопоставления семантико-синтаксических формул абзацев и семантико-синтаксических формул текстов. Выведены канонические модели абзацев и обобщенные семантико-синтаксические модели текстов двух жанровых разновидностей технических описаний английского научно-технического подъязыка "вычислительная техника". Полученные результаты дают новые эксплицитные критерии для различения жанров и типологизации текстов. Проведенное исследование ориентировано на последующую компьютерную реализацию. Предложенная методика может найти применение при описании семантико-синтаксической структуры текстов любого английского научно-технического подъязыка.

Р Э З Ю М Э

КУНЦЭВІЧ СВЯТЛАНА ЯЎГЕНАЎНА

СЕМАНТЫКА-СІНТАКСІЧНАЯ АРГАНІЗАЦЫЯ ТЭКСТАЎ ТЭХНІЧНЫХ АПІСАННЯЎ (НА МАТЭРЫЯЛЕ АНГЛІЙСКАЙ ПАДМОВЫ "ВЫЛІЧАЛЬНАЯ ТЭХНІКА")

Навукова-тэхнічны тэкст, жанр, тэхнічнае апісанне, змест тэксту, логіка-семантычная структура тэксту, элементарны кампанент зместу абзаца, семантыка-сінтаксічная мова для фармальнага прадстаўлення зместу тэксту, асноўны змест тэксту, факультатыўны змест тэксту, правілы перабудовы семантыка-сінтаксічных формул абзацаў, кананічная мадэль абзаца, абагульненая семантыка-сінтаксічная мадэль тэксту.

Даследаванне выканана з мэтай праверкі гіпотэзы адносна існавання для тэкстаў дастаткова вузкай прадметнай галіны і аднолькавай жанравай прыналежнасці канчатковага набору іх абагульненых семантыка-сінтаксічных мадэлей. Распрацаваны новы комплексны падыход да апісання зместу навукова-тэхнічных тэкстаў, які прадугледжвае тры ўзроўні іх прадстаўлення: логіка-семантычны, тэматычны і семантыка-сінтаксічны. Прапанаваны спосаб здабывання асноўнай і факультатыўнай інфармацыі з навукова-тэхнічнага тэксту. Уведзена новая адзінка аналізу навукова-тэхнічнага тэксту -- элементарны кампанент зместу (ЭКЗ) абзаца. Створана новая семантыка-сінтаксічная мова для фармальнага прадстаўлення зместу англійскага навукова-тэхнічнага тэксту. Апісана методыка стварэння і супастаўлення семантыка-сінтаксічных формул абзацаў і семантыка-сінтаксічных формул тэкстаў. Выведзены асноўныя кананічныя мадэлі абзацаў і абагульненыя семантыка-сінтаксічныя мадэлі тэкстаў дзвюх жанравых разнавіднасцей тэхнічных апісанняў англійскай навукова-тэхнічнай падмовы "вылічальная тэхніка". Атрыманыя вынікі даюць новыя экспліцытныя крытэрыі для разлічэння жанраў і тыпалагізацыі тэкстаў. Выкананае даследаванне арыентавана на далейшую камп'ютэрную рэалізацыю. Прапанаваная методыка можа знайсці прымяненне пры апісанні семантыка-сінтаксічнай структуры тэкстаў любой англійскай навукова-тэхнічнай падмовы.

S U M M A R Y

KUNTSEVITCH SVETLANA EVGENIEVNA

SEMANTIC-SYNTACTICAL STRUCTURE OF TWO TEXT VARIETIES OF TECHNICAL DESCRIPTION (on the basis of the English sublanguage of computer science)

Scientific-technical text, genre, technical description, text content, logical-semantic text structure, thematic text structure, elementary content component of the paragraph , semantic-syntactical code for formal presentation of text content, main text content, optional text content, rules of semantic-syntactical transformations of paragraph formulas, canonical paragraph model, generalized semantic-syntactical text model.

This scientific investigation is fulfilled for the purpose of verifying the hypothesis that one can build a limited number of generalized semantic-syntactical models for texts belonging to one genre of the same scientific-technical sublanguage. A new complex method of text content description is worked out. This method presupposes three levels of text presentation: logical-semantic, thematic and semantic-syntactical. A way of extracting essential and optional text information is described. A new unit for text analysis is suggested -- an elementary content component (ECC) of the paragraph . A special semantic-syntactical code for formal presentation of text content has been elaborated. Methods of creating and comparing semantic-syntactical paragraph and text formulas are worked out. Certain canonical paragraph models and generalized semantic-syntactical text models are deduced. The results obtained give new explicit criteria for genre differentiation. The methods suggested can be applied for the description of the semantic-syntactical text structure of any English scientific-technical sublanguage. The investigation itself has a direct computer orientation.

Научное издание

КУНЦЕВИЧ Светлана Евгеньевна

СЕМАНТИКО-СИНТАКСИЧЕСКАЯ ОРГАНИЗАЦИЯ ТЕКСТОВ ТЕХНИЧЕСКИХ ОПИСАНИЙ

(на материале английского подъязыка "вычислительная техника")

Специальность 10.02.19 - общее языкознание

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата филологических наук

191(6825)

Ответственный з
Подписано в печ
Формат бумаги 6
Усл. печ. л. 1,09.
Тираж 80 экз. За
Минский государ
220062, г.Минск,
Полиграфически
220062, г.Минск,

ерситета.