

0-784473

На правах рукописи



**ФААЛ-ХАМЕДАНЧИ Марьям**

**ТЕЗАУРУСНОЕ ПРЕДСТАВЛЕНИЕ  
ТЕРМИНОВ НЕФТЕГАЗОВОЙ ОТРАСЛИ  
В РУССКОМ И ПЕРСИДСКОМ ЯЗЫКАХ**

Специальность 10.02.20 – сравнительно-историческое, типологическое и сопоставительное языкознание

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата филологических наук

**Москва – 2010**

Работа выполнена на кафедре общего и русского языкознания  
филологического факультета Российского университета дружбы народов

**Научный руководитель:**

доктор филологических наук, профессор Красина Елена Александровна

**Официальные оппоненты:**

доктор филологических наук, профессор Трофимова Галина Николаевна,  
*Российский университет дружбы народов*

кандидат филологических наук, профессор Каменева Марина Самойловна,  
*Институт востоковедения РАН*

**Ведущая организация:** **Международный славянский институт**

Защита диссертации состоится 28 мая 2010 года в 15.00 часов  
на заседании диссертационного совета Д 212.203.12  
при Российском университете дружбы народов  
по адресу: 117198, Москва, ул. Миклухо-Макляя, 6, ауд. 436

С диссертацией можно ознакомиться в Учебно-научном информационном  
центре (Научной библиотеке) РУДН по адресу: 117198, Москва, ул.  
Миклухо-Макляя, 6, ауд. 436

Автореферат диссертации размещен на сайте РУДН – [www.rudn.ru](http://www.rudn.ru)

Автореферат разослан 26 апреля 2010 г.

Ученый секретарь  
диссертационного совета  
кандидат филологических наук, доцент

НАУЧНАЯ БИБЛИОТЕКА КГУ



0000610423

Н.Ю. Нелובה

### **Общая характеристика работы**

Диссертация посвящена разработке принципов представления терминов нефтегазовой отрасли в двуязычном русско-персидском и персидско-русском словаре тезаурусного типа с перспективой создания переводного отраслевого словаря тезаурусного типа.

**Объектом исследования** послужили термины и терминологические сочетания нефтегазовой отрасли в русском и персидском языках, которые образуют фрагмент лексической системы в каждом из языков и при использовании информационно-поискового языка создают основу отраслевого словаря тезаурусного типа.

**Предметом исследования** явились способы и приемы сопоставительного лексикографического представления терминов нефтегазовой отрасли на основании использования информационно-поискового языка и применении информационных технологий в двуязычном переводном словаре тезаурусного типа. Сопоставительный аспект способствовал не просто выявлению сходств и различий в терминосистемах русского и персидского языков, но и выбору релевантных приемов и способов словарного представления терминов в этих разноструктурных языках.

**Актуальность исследования** определяется рядом факторов. Во-первых, отраслевые переводные словари становятся все более востребованными как для прикладных целей, в частности, ведения совместного бизнеса или производства, так и для теоретических и практических лексикографических разработок и создания многоязычных лингвистических баз данных. Во-вторых, нефтегазовая терминология является основополагающей в сфере сотрудничества нефтедобывающих стран, а поэтому она является важным показательным фрагментом языковой системы с точки зрения интернационализации функционирования и развития лексических ресурсов языка. Наконец, в связи с появлением новых технологий, формированием предметных и лингвистических электронных баз данных, создание отраслевых электронных словарей тезаурусного типа требует разработки новых методов и методик изучения и описания лексического и лексикографического материала.

**Целью** данной работы является сопоставительное исследование терминосистем нефтегазовой отрасли в русском и персидском языках с перспективой разработки принципов и приемов создания двуязычного переводного словаря тезаурусного типа.

Поставленная цель требует решения ряда конкретных задач:

- описание сходств и различий между тезаурусными словарями различных типов: общеязыкового, отраслевого и переводного тезауруса;
- на основе сопоставительного анализа языкового материала разработать корпус терминов нефтегазовой отрасли для лексикографического представления;
- отобрать способы лексикографического описания терминов и обработки информационно-поискового языка для тезаурусного представления корпуса терминов нефтегазовой отрасли в сравниваемых языках;

- представить термины нефтегазовой отрасли как элементы базы данных и как единицы тезауруса;
- разработка принципов представления терминов в двуязычном словаре тезаурусного типа и описать его структуру.

**Основные положения, выносимые на защиту:**

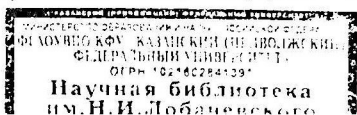
1. Сопоставительный анализ терминов нефтегазовой отрасли в разноструктурных русском и персидском языках позволяет уточнить корпус базы данных и лексикографических источников данной терминосистемы.
2. Одновременно выступая в качестве элементов отраслевой базы данных и словаря-тезауруса, термины нефтегазовой отрасли образуют лексическую микросистему, которая становится основой двуязычного – русско-персидского и персидско-русского переводного словаря тезаурусного типа.
3. Информационно-поисковый язык, который лежит в основе разработки словарей-тезаурусов, имеет универсальный характер, но нуждается в некоторой адаптации с учетом сходств и различий сопоставляемых языков.
4. Информационные технологии позволяют создавать максимально сложные лексикографические источники на основании информационно поискового языка определенного типа. Переводные дву- и многоязычные словари тезаурусного типа, соответственно, удваивают и умножают количество указателей и списков-входов в них.
5. Сложная структура словаря требует комплексной методики его разработки и создания.

**Материалом исследования** прослужили общеязыковые и специальные словари русского и персидского языков, языковые и отраслевые базы электронные данных; информационные материалы и тексты нефтегазовой тематики на сопоставляемых языках.

**Научная новизна** работы состоит в том, что:

- впервые создается корпус языкового материала для нефтегазовой отрасли и разрабатываются принципы представления терминосистем русского и персидского языков в двуязычном переводном словаре тезаурусного типа;
- применяется комплексная методика использования информационно-поискового языка для наиболее полного представления заголовочного слова и его толкования в электронной версии отраслевого словаря-тезауруса;
- сведение в единый источник терминов двух неблизкородственных и разноструктурных языков востребовано практической деятельностью в нефтегазовой отрасли для двух нефтедобывающих стран и, являясь насущной задачей времени, позволяет решать ряд прикладных задач не только в области сопоставительной лексикографии и разработки специальных баз данных, но и активно способствует сотрудничеству в области бизнеса и индустриальной деятельности.

**Теоретической базой** диссертации послужили труды, заложившие основу и развивающие исследование терминосистем в аспекте создания баз данных и словарей тезаурусного типа: Азарова И.В., 2003; Апресян Ю.Д., 1995; Берков В.П., 2004; Караулов Ю.Н., 1982; Лаврѐнова О.А., 2001;



Морковкин В.В., 1970; Шемакин Ю.И., 1974; Atkins S., 2003; Croft W., 2004; Fillmore C.J., 2003; Halliday M.A.K., 2004 и др.

**Методологической базой** диссертационной работы стал комплексный анализ лексических единиц с терминологическим значением с целью создания двуязычного переводного словаря тезаурусного типа. Наряду с общенаучными методами (индукция, дедукция, наблюдение, обобщение, абстрагирование) используются сопоставительно-типологический метод, описательный метод при характеристике особенностей понятийной и языковой семантики, логико-семантический метод при выявлении содержательной сущности категорий и понятий, а также лексикографические стратегии построения ассоциативных и идеографических словарей и баз данных, включая информационные технологии. (см. Ю.Н. Караулов, О.А. Лаврёнова, В.В. Морковкин, J. Aitchinson, W.Croft и др.).

**Теоретическая значимость** исследования состоит в том, что сформулированные в работе положения и выводы развивают, уточняют и дополняют существующие принципы разработки лингвистических баз данных и их представление в словарях тезаурусного типа; в разработке принципов словаря нового типа – двуязычного переводного отраслевого словаря-тезауруса по нефтегазовой тематике; в разработке принципов и методики составления двуязычного электронного тезауруса с опорой на передовой лексикографический опыт; в сопоставлении двух неблизкородственных и разноструктурных языков на лексическом и лексикографическом материале; в опоре на типологически релевантные характеристики двух сопоставляемых языков – русского и персидского. Содержащиеся в работе положения и выводы имеют значение для дальнейших теоретических разработок в области систематизации языкового лексического материала и его лексикографического представления в словарях тезаурусного типа.

**Практическое значение** работы определяется возможностью использования ее результатов в лексикографической практике при составлении дву- и многоязычных словарей тезаурусного типа, особенно тематических или отраслевых словарей и, соответственно, созданию специальных лингвистических баз данных. Сопоставительный аспект диссертации свидетельствует о том, что ее результаты могут использоваться в практике преподавания ряда лингвистических дисциплин: сопоставительной лексикологии и грамматики, сопоставительной типологии, теории перевода; а также спецкурсов по сопоставительной лингвистике, в практике составления персидско-русских и русско-персидских словарей, при написании квалификационных научных и научно-практических работ.

**Объективность и достоверность основных положений и выводов** обеспечивается привлечением сведений из различных областей гуманитарного знания, обширным корпусом трудов по лингвистике, применением комплексной методики изучения языкового материала и многочисленными аутентичными источниками языкового материала.

**Апробация работы.** Основные результаты работы обсуждались на заседаниях кафедры общего и русского языкознания филологического

факультета РУДН, на международном семинаре Lexisom (Барселона, 2009). По теме диссертационного исследования опубликовано 3 статьи, в том числе 2 статьи опубликованы в научных рецензируемых изданиях, рекомендованных ВАК.

**Структура диссертации.** Работа состоит из введения, четырех глав, заключения, списка литературы и источников, включающего более 200 наименований, и приложений.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Глава 1. Введение в теорию и практику лексикографии** содержит основные вопросы лексикографии и рассматривает новые направления в лексикографии.

В первой части обсуждается основная задача лексикографии, это - объединять теорию и практику: она представляет собой научную разработку принципов и методов словарного описания лексики, а также составление всевозможных словарей, которые отражают различные аспекты существования и функционирования слов в системе языка.

Далее, отмечается, что история развития лексикографии свидетельствует о ее прочных связях с теорией языка и, в первую очередь, с лексической семантикой. Лингвистика XIX в. вовсе не интересовалась вопросами теории лексикографии, но для лингвистики XX в. характерно стремительное развитие лингвистической семантики и лексикографии, которое отражается в работах таких российских и зарубежных языковедов, как Л.В. Щерба, Ш. Балли, Э. Сэпир, К. Эрдман, Дж. Фирт, В.В. Виноградов, Ю.Н. Караулов, Л.А. Новиков и др.

Предмет современной лексикографии рассматривается в следующей части и объясняется, что он связан с изучением словарного состава языка как сложной лексической системы, в которой слова – элементы характеризуются содержательно через отношения к другим словам, т.е. структурируются по месту в данной лексической системе.

При этом слово рассматривается в совокупности всех своих свойств, так как значение слова определяется не только его соответствием понятию, но и обусловлено принадлежностью к определенной части речи, реализацией грамматических категорий. Оно также зависит от контекстов употребления, от семантических связей с другими словами, от семантического соотношения этого слова с синонимами и вообще с близкими по значениям и оттенкам значений словами, а также от экспрессивной и стилистической окраски слова. Границы слова и его значения определяются только на фоне лексико-семантической системы языка в целом. Семантика изучает систему значения слова, выделяет его особенности и характеристики, на основе которых лексикография может выполнять свои задачи. Семантика также выделяет те законы и константы, которые действуют во всех языках и таким образом дает лексикографии универсальную основу. Например, специальная лексика характеризует научно-технические тексты, а далее, в научно-техническом контексте лексика отражает характеристики, которые отличают слова как

специальные термины, наконец, выделяется терминология как особая область лексикологии и лексикографии. Таким образом, терминология является предметом технической лексикографии и изучает характеристики и типы специальной лексики.

Далее резюмируются используемые различными словарями подходы представления аспектов и лексической единицы и отмечается, что словари используют различные подходы для отражения разных аспектов лексической единицы. В толковых словарях, например, лексическое значение описывается как единое целое, без четкого выделения таких его аспектов, как сигнификативный, структурный, парадигматический, синтагматический, прагматический, денотативный. Однако перечисленные аспекты имплицитно содержатся как в самих словарных статьях, так и в иллюстрациях: например, парадигматический аспект лексического значения представлен синонимами, антонимами, словами той же лексико-семантической группы, в которую входит данное слово, а синтагматический аспект реализуется в типичных словосочетаниях. Денотативный аспект содержания представлен непосредственно как список значений в словарной статье, а на сигнификативный аспект указывается в том случае, если слово обозначает конкретный предмет. Прагматический же аспект, как правило, обнаруживается в стилистических пометах.

В переводных словарях вместо толкования значений на одном и том же языке даются переводы этих значений на другой язык, например: *газификатор* – *gasifier*; *бурение* – *drilling*.

В идеографических словарях центральным принципом организации словаря избирается понятийное содержание. Здесь представлена связь понятия и слова в направлении от понятия к слову, а слова размещаются по принципу их смысловой близости, чтобы облегчить поиск необходимого значения.

Вопрос о том, какие аспекты лексической единицы должны быть представлены в словаре, и о выборе способов их представления решается в зависимости от целей и задач словаря с учетом фактора пользователя словаря. Алфавитный и идеографический подходы к представлению слова в словарях опираются на различные способы, соответственно предмету разработки. Если слово является двусторонним единством, соединяющим в себе форму и значение, то алфавитный подход реализуется в направлении от формы к значению (от языкового знака к концепту), а идеографический – от значения к форме (от концепта к языковому знаку). Однако оба подхода зависят друг от друга. Так, при толковании значения лексической единицы используются тождественные по значению слова-синонимы, родовидовые связи слова или отрицательные определения. Идеографический подход не совсем свободен от систематического подхода, и алфавитный порядок расположения единиц используется либо внутри идеографических разделов, либо в виде приложений, т.е. присутствует как неотъемлемая часть лексикографического источника.

В последней части обсуждаются современные лексикографические характеристики и отмечается, что в настоящее время в лексикографии в процессе составления словарей активно применяются новые

лексикографические технологии. В частности, применение компьютерных технологий позволяет лексикографу расположить материал в любом, необходимом ему порядке, определить частотность слова, выполнить рисунки, найти иллюстрирующие цитаты, привести этимологию слова и т.д. Появление корпусной лексикографии способствует не только функциональной деятельности лексикографии, но и оперативному контролю над развитием и изменением словарного состава языка, да и языка в целом.

Применение в лексикографии современных технологий способствует развитию исследования различных аспектов языка и раскрытию новых универсальных подходов к составлению словаря. Так, до недавнего времени в лексикографии существовало два основных способа фиксации лексических единиц в словарях: алфавитный и идеографический, однако в последнее десятилетие развивались новые направления в области лексикографии, как например, электронные словари, корпусная лексикография и интернет-тезаурус, типа Multilex, WordNet и FrameNet, которые позволяют постоянно увеличивать количество лексических единиц, фильтровать и упорядочивать словарный материал. Создание корпусов текстов как лингвистических баз данных является серьезным шагом на пути к созданию словарей нового поколения и решению многих проблем современной лексикографии.

Киберлексикография, или лексикография с использованием информационных технологий, с одной стороны, позволяет пользователям быстро и эффективно осуществлять поиск нужной информации при обращении к разнообразным лексикографическим источникам, а с другой стороны – дает лексикографу возможность изучать и анализировать способы человеческого мышления в отношении к поиску информации.

Быстрое развитие и всемирное распространение различных лексикографических подходов на основе новейших информационных технологий обеспечивает наибольшую адаптацию лексикографических приемов обработки информации к общечеловеческому способу мышления и созданию обширных языковых баз данных.

**Глава 2. Тезаурус как идеографический словарь особого типа** описывает и общую характеристику тезауруса, и его специфику.

Обсуждая мнения различных лингвистов о тезаурусе, мы приходим к выводу, что тезаурусы – это, в определенном смысле, «обращенные толковые словари», по Ю.Н. Караулову и др. Однако у тезаурусов еще есть особенные черты, которые отличают их от других лексикографических источников. В отличие от толковых словарей, которые представляют только один «вход», т.е. заглавное слово, предшествующее толкованию, ряду синонимов, переводу и т.д., стандартный тезаурус имеет, как правило, четыре вход-указателя:

1) **К→К** – вход от концепта к концепту, от одного понятия к другому, воплощенный в схеме вертикально и горизонтально связанных друг с другом понятий, т.е. понятий, связанных родовидовыми, видо-видовыми и другими типами корреляций.

2) **К→З** – вход от концепта к знаку, от понятия к слову – основная часть тезауруса.

3) З→К – вход от знака к концепту, соответствующий алфавитному списку слов с указанием их адресов в понятийных полях тезауруса.

4) З→З – вход от знака к знаку, необходимый для тех тезаурусов, в которые, наряду со словами, входят и словосочетания, и в связи с этим от слов, входящих в словосочетание, дается отсылка к отдельным словам, например от слова *автоматический* к слову *автоматизация*.

Далее отмечается, что не все тезаурусы включают все эти четыре входа. Данная структура встречается, в большинстве случаев, у Информационно-поискового тезауруса (далее – ИПТ). Такой подход трудно осуществить в целом, особенно при составлении общезыковых тезаурусов, поскольку общезыковые тезаурусы, как правило, используют естественный национальный язык. Естественный язык обладает рядом свойств, осложняющих его использование для записи и поиска информации. Точное значение многих слов можно определить только исходя из контекста, в котором они употреблены. При составлении ИПТ создается информационно-поисковый язык, или ИПЯ, который является искусственным языком, предназначенным для описания основного смыслового содержания текстов (документов) или их частей, а также для выражения смыслового содержания информационных запросов с целью реализации информационного поиска.

Переход от понятия к слову, от концепта к языковому знаку является отличительной чертой тезауруса, в которой сосредоточиваются все его языковые ценности и которая отражает систематизированную характеристику человеческого мышления. Кроме того, упорядоченность двух родов – систематическая и алфавитная – отличает этот словарь от толкового и определяет главное структурное его свойство – наличие более чем одного входа-указателя. Таким образом, тезаурус расширяет число возможных точек доступа к словарному составу языка, тем самым расширяя семантическое пространство лексики естественного языка.

Технология тезаурусного описания суммируется в следующем параграфе. Она включает процедуры формирования списков слов-понятий и слов-отношений и их группировки в парадигматические классы. Таким образом, тезаурус является средством описания парадигматической структуры плана содержания языка, потому что в тезаурусе наглядно отражается факт наличия парадигматической семантической связи между лексическими единицами путем объединения единиц в поля с общим смысловым признаком, чем и определяется место каждой лексемы как единицы системно организованного смыслового пространства.

С учетом того, что характеристики информационно – поисковых (ИПТ) тезаурусов по большей части подходят к техническому переводному тезаурусу, в следующем параграфе рассматриваются их структуры и отмечается, что в ИПТ обычно фиксируются следующие парадигматические отношения и связи:

- «род – вид» или другое иерархическое отношение («выше – ниже», «шире – уже»);
- отношение синонимии (условной эквивалентности);
- ассоциативные отношения различного вида.

Родовидовое отношение существует на уровне понятия, а не лексики, т.е. связь рассматривается между двумя понятиями, и она существует независимо от языка. Например, «*каротаж*», «*logging*», «*چاه پیمایی*» – разные знаки для представления одного и того же понятия, которое является видом нефтяной технологии.

Для нефтегазовой терминологии можно определить следующие основные родовидовые отношения:

1). **Виды материалов и продукции:** разделяются по происхождению, составу и функции. Например, *нефтепродукт: бензин; газойль; мазут.*

2). **Процессы, операции и технологии** классифицируются на основе способа выполнения, объекта действия, используемого оборудования, цели действия, характеристики действия и т.д. Например, *крекинг: коксование; термический крекинг, газификация: газификация нефти; газификация угля.*

3). **Оборудование, инструменты, приспособления.** Например: *якорь: газовый якорь; якорь перфорированного хвостовика; якорь подвесного типа.*

В естественном языке синонимия определяется как отношение между двумя словами, которые называют одну и ту же вещь, но соотносят ее с разными понятиями и, тем самым, через название, вскрывают различные свойства данной вещи. Отмечаются четыре основных типа синонимических отношений:

1. полные синонимы (*аванс – предоплата*);
2. синтаксические синонимы (*авария на транспорте – транспортная авария*);
3. квазисинонимы (*сила гравитации – сила тяжести*);
4. словообразовательные синонимы (*природоохранительный – природоохранный*).

В двуязычных тезаурусах представление синонимического отношения является наиболее сложной задачей: далеко не все синонимические варианты лексики единицы имеют эквиваленты в другом языке. Однако тезаурус обеспечивает достаточно надежные способы для представления значения синонимических вариантов.

Ассоциативные отношения устанавливаются между ключевыми словами (далее – КС), принадлежащими к одной и той же или даже разным смысловым категориям и произвольным уровням иерархии. Они аналогичны ассоциативным связям, возникающим в сознании человека, когда представление об одном объекте вызывает представления о других, например, *аварийная ситуация: авария, ошибочное действие, пожар, нарушение правил, ликвидация аварии, аварийно-спасательные службы, безопасность* и т. д. Ассоциативные связи в тезаурусе, как правило, семантически не различаются. Однако используются следующие виды ассоциаций: «целое – часть», «причина – следствие», «близость в пространстве или во времени», «антонимия», «предмет – обычная область его применения» и т. д.

Далее суммируются указатели тезауруса. Информационно-поисковый тезаурус (ИПТ) обычно включает две основные части:

- указатель иерархических отношений;

- алфавитный указатель.

Дополнительно может быть построен пермутационный указатель, т.е. указатель, в котором в алфавитном порядке перечислены все отдельные слова, входящие как компоненты словосочетаний, обозначающих дескрипторы, и для каждого из них указаны все дескрипторы, в состав которых входят эти слова.

Алфавитный указатель представляет собой алфавитный перечень всех ключевых слов и дескрипторов, включенных в тезаурус, с указанием парадигматических отношений между ними. Заголовочной единицей тезаурусной статьи данного указателя является дескриптор. Все остальные части класса условленного дескрипторной эквивалентности считаются условными синонимами этого дескриптора и связаны с ним при помощи отсылочных помет. Указатель иерархических отношений содержит полные иерархические деревья, отражающие все видовые дескрипторы для каждого родового дескриптора, и ступени получения видовых дескрипторов. Иерархические деревья в виде так называемой точечной записи располагаются в этом указателе по семантическим категориям.

Пермутационный указатель терминов является дополнительным входом в тезаурус, к которому обращаются, когда нужные термины не удается найти непосредственно в разделе «Лексико-семантическое собрание терминов». Здесь каждое КС, состоящее из двух или более слов, т.е. представленное сочетаниями слов, включается под всеми составляющими столько раз, сколько в нем составляющих слов. Поиск термина производится по наиболее семантически важному слову, а в случае отсутствия такового – по любому значащему слову термина-сочетания. Каждый термин встречается в пермутационном указателе в алфавитном порядке столько раз, сколько значащих слов он содержит.

В продолжении отмечается, что представление взаимосвязанной, или синоптической схемы понятий является одной из отличительных особенностей тезауруса. Фиксируя семантические отношения между составляющими его единицами, тезаурус пытается отразить модель человеческого мышления и способы осмысления мира человеком. Зависимая от области применения тезауруса, его синоптическая различается, т. к., с одной стороны, отдельные области знания развиваются неравномерно, а с другой стороны, обобщению в единую и целостную систему препятствуют нерешенные проблемы синтеза специфических картин мира.

Используя различные указатели-входы, тезаурус позволяет получить нужную информацию различными способами. При этом тезаурус опирается на достижения лексикографии, лингвистической семантики, онтологии, культурологии и др. Так, иерархический указатель в первую очередь отражает онтологическую связь между понятиями, а алфавитный указатель раскрывает лексико-семантическую связь между языковыми единицами.

В последней части автор приходит к выводу, что два основных типа тезаурусов – это общезыковые и информационно-поисковые. Общая структура тезауруса определяется на основе его главной задачи и схемы, хотя по структуре принципиальной разницы между общезыковым и информационно-поисковым тезаурусом нет. Общезыковые тезаурусы основываются на

логической рубрикации всего понятийного содержания лексики, а в терминологических отраслевых словарях такая детализация дескрипторов становится невозможной, т. к. термины каждой области науки и технологии приобретают свое особенное значение исключительно в рамках данной отрасли. При этом в последние десятилетие тезаурусы разрабатывались практически для всех отраслей деятельности, и было создано достаточно большое количество специализированных тематических тезаурусов.

**Глава 3. Принципы составления двуязычного отраслевого тезауруса** раскрывает этапы создания тезауруса, описывает методики отбора и обработки ключевых слов и процедуры выявления парадигматических связей.

Описание работы по формированию многоязычных тезаурусов определяется наличием или отсутствием одноязычных тезаурусов по определенной тематике на языках-компонентах. При отсутствии одноязычных тезаурусов по заданной тематике на языках-компонентах, как правило, сначала создают одну одноязычную версию и на ее основе строят остальные. Возможна также совместная разработка одновременно всех или нескольких одноязычных версий при взаимном согласовании состава дескрипторов и их парадигматических связей.

Далее резюмируется работа по созданию тезауруса в следующие основные этапы:

- определение тематического охвата разрабатываемого тезауруса;
- отбор словарного состава для отражения тематики данной области знания и формирование словника тезауруса;
- составление классификационных схем понятий данной области знания;
- совмещение словника и классификационных схем понятий и формирование на их основе иерархической части тезауруса;
- построение алфавитной (лексико-семантической) и других частей тезауруса;
- проведение экспериментального определения с доработкой и расширением состава первой редакции тезауруса;
- разработка системы дальнейшего пополнения тезауруса, т.е. правил и порядка внесения в него изменений и дополнений в процессе его использования.

Среди методов составления тезаурусов наиболее традиционным является эмпирический метод, т.е. отбор терминов путем обработки текстовых документов по данной тематике. Обработка включает в себя изучение содержания документов, выделение из них ключевых слов и словосочетаний и регистрацию каждого ключевого слова в стандартной форме на карточках. Карточки расставляются в алфавитном порядке. Таким образом, получается алфавитный словник тезауруса.

В следующей части отмечается, что целесообразность включения ключевых слов в тезаурус следует оценивать по нескольким факторам, которые тесно взаимосвязаны:

- по точности и приемлемости данного КС с точки зрения терминологии рассматриваемой области науки или техники;
- по его связям с КС, отобранными ранее;
- по решению специалистов в данной области знания.

Затем проводится систематизация словарика тезауруса и взаимное дополнение классификационных схем и словарика. На этом этапе выделяются и синонимы, которые объединяются в группы условной эквивалентности. В каждой группе условной эквивалентности выбирается наиболее общий и употребительный термин, т.е. дескриптор, который будет представлять эту группу. В окончательных схемах, совокупность которых составляет иерархический указатель тезауруса, представлены только дескрипторы.

Типы лексических единиц, которые включаются в словарь тезауруса суммируются в следующие:

- одиночные существительные, прилагательные, наречия, некоторые причастия, очень редко, глаголы;
- именные словосочетания;
- лексически значимые компоненты сложных слов;
- сокращения слов и словосочетаний.

Вопрос о разделении или сохранении словосочетаний и сложных слов является одной из центральных проблем любой лексикографической работы. В общеязыковых словарях сходный вопрос существует для устойчивых выражений и оборотов речи, а решение вопроса заключается в характеристиках данных словосочетаний и в различиях устойчивых выражений и оборотов речи. Тогда в технических и отраслевых словарях, поскольку лексические единицы являются терминами, словосочетания должны различаться на основе одиночных понятий.

Далее отмечается, что составление тезауруса на основе категориальной синоптической схемы, позволяет быстро и актуально создать иерархические деревья по отношению «род-вид». Иерархические деревья выстраиваются по отношению «род-вид» и распределяют ключевые слова по более узким семантическим категориям, в конечных точках которых получаются эквивалентные слова и словосочетания. Среди этих слов и словосочетаний выбирается дескриптор как их категориальный представитель.

Иерархические деревья должны располагаться по основным семантическим категориям тематических разделов данной области деятельности. Наиболее значительным этапом в составлении тезауруса является выбор или разработка способа классификации схемы понятий. Все дескрипторы тезауруса должны предстать распределёнными по семантическим категориям, которые представляют собой некоторые наиболее общие понятия соответствующей области знания и входят в тезаурус в качестве родовых для большой группы дескрипторов. Главные семантические категории определяются специально для каждого тематического раздела на основе анализа области деятельности. Например, в **Exploration and Production Thesaurus** выделяется десять семантических категорий: (1) **Common attribute** – *общий признак*; (2) **Earth and space concepts** – *понятия, связаны с землей и*

*космосом*; (3) **Economic factor** – *экономический фактор*; (4) **Equipment** – *оборудование*; (5) **Material** – *материал*; (6) **Operating condition** – *рабочее условие*; (7) **Organism** – *организм*; (8) **Phenomenon** – *явление*; (9) **Process** – *процесс*; (10) **Property** – *свойство*.

Процесс распределения ключевых слов по категориям не зависит от их принадлежности к частям речи и носит исключительно семантический характер. Так, имена существительные могут относиться к любой из главных семантических категорий. Принадлежность прилагательного, в синтагматическом плане (в плане текстовых отношений) выражающего характеристику, в парадигматическом плане (т. е. вне зависимости от текста) следует определять по тому существительному, которому оно условно эквивалентно в данной области знания. Например, *КС сварочный* = для *сварки* (категория **Процесс**), но *КС сварной* не эквивалентно существительному *сварка* и относится к категории **Свойство**. Эквивалентность устанавливается между ключевыми словами, если замена одного из них на другое не влияет на результаты поиска информации в одном или нескольких языках.

В последней части рассматриваются ассоциативные отношения. Ассоциативные отношения выявляются путем семантического анализа дескрипторов тезауруса. Решения об использовании тех или иных ассоциативных отношений должны приниматься на основании специфики лексического материала и с учетом задачи поиска информации. Это значит, ссылка на ассоциативный дескриптор должна приводиться в том случае, когда запрос данного дескриптора может потребовать, также, поиска документов, содержащих дескрипторы, подобные данному. Наиболее эффективные методы разработки ассоциативных связей – это ассоциативный эксперимент, анализ текстов, документов и специальное исследование процесса поиска пользователями информации. Поскольку в отраслевой тезаурус включаются термины отдельной области, то с прагматической точки зрения в таком тезаурусе наиболее значительными оказываются ассоциативные отношения.

**Глава 4. Двуязычный переводный тезаурус нефтегазовой отрасли** описывает сходства и различия информационно-поискового и технического переводного тезаурусов, основные характеристики предлагаемого отраслевого переводного тезауруса, а также ассоциативный эксперимент, который сделан автором с целью проверки итогов работы.

В первой части отмечается, что хотя дву- и многоязычные ИПТ способны по-разному представлять эквиваленты определенной лексической единицы в других языках, они не являются в полной мере адекватными, чтобы использоваться в качестве переводного тезауруса. Их недостатки заключаются в том, что на практике бывает сложно представить существующие точные эквиваленты определенного термина в других языках. Кроме того, в них игнорируются некоторые языковые аспекты лексической единицы, которые существенно помогают пользователям тезауруса в поиске языковой информации. Исходя из этого, отметим, что предлагаемый отраслевой переводный тезаурус создается на основе смешенной схемы ИПТ и тематического словаря.

Главная цель переводного отраслевого тезауруса суммируется в следующий и отмечается, что он являет собой представление эквивалента технического термина в выходном языке таким, чтобы пользователи тезауруса без особых усилий могли подобрать самый подходящий эквивалент среди множества возможных эквивалентов. Потом обсуждается, что понятийные значения эквивалентов термина в разных языках не всегда полностью совместимы, а между эквивалентами в разных языках существуют следующие типы связей: 1) **Полная эквивалентность**: значение слова и его эквивалента в другом языке практически полностью совпадают. Например, *physics/ физика/ فيزيك (fizik)*. 2) **Неполная эквивалентность**: значения слов не совпадают, но пересекаются. Например, *environment/ окружающая среда/ محیط زیست (mohit-e zist)*. 3) **Частичная эквивалентность**: значение одного слова шире, чем значение другого, как правило, аналогично родовидовому отношению «шире – уже». Например, *document/ документ/ مدارک (madrek)*, где значение документ на русском языке шире, чем значение в персидском языке مدرک (*madrak*), передаваемое формой множественного числа مدارک (*madrek*) ‘документы’. 4) **Эквивалентность «один ко многим»**: значение одного слова соответствует совокупности значений нескольких слов. Например, *broadcasting/ радиопередача; телепередача/ رادیو و تلویزیون /radiyu va televisiyun/*. 5) **Неэквивалентность как отсутствие эквивалентов**: не существует эквивалентов слова в выходном языке. Например, *upstream/ звенья нефтяной отрасли, связанные с разведкой и добычей нефти/ بالادستی (baladasti)*. Информационный поисковый тезаурус использует разные способы для решения данной проблемы, в том числе искусственный информационный язык, который не отвечает требованиям пользователей переводного тезауруса. Таким образом, переводный тезаурус вынужден использовать и лексикографические приемы переводных словарей. Но и переводные словари далеко не полностью удовлетворяют потребности пользователей, т.к. из-за особенностей процесса возникновения нового термина специальная терминология богата различными видами синонимии и полисемии, что противоречит общепринятому мнению об отсутствии синонимии и полисемии в терминах. Из-за богатства грамматической структуры, особенно в русском языке, широко распространены синтаксические синонимы и номинативные варианты в их употреблении. Таким образом, один и тот же термин может подключать различные номинативные варианты и, следовательно, иметь полные и частные синонимические эквиваленты, которые относятся к одному понятию. Например, англ. *associated gas*, в близких областях имеет различные синонимы: *accompanying gas; wellhead gas; dissolved gas; oil gas; oil-well gas; petroleum gas; trip gas*; Следовательно, имеет и различные русские синонимические эквиваленты: *путный газ; свободный газ; нефтяной газ; газ из добывающей скважины; нефтяной промысловый газ; растворённый в нефти газ; газ добывающей скважины; газ газовой шапки; путный нефтяной газ*. Когда речь идет о совпадении русского и персидского эквивалентов, данная проблема осложняется, т.к. английский язык становится переходным языком, и к осложнению проблемы приводится список несовпадений персидских

эквивалентов, с одной стороны, и русских – с другой, а также используется специальная пометка (пояснение). Такой подход помогает понимать значения термина в выходном языке и до некоторой степени обеспечивает правильное употребление синонимов. Одновременно используя способы тематических словарей и информационно-поисковых тезаурусов, переводный тезаурус обеспечивает еще более вероятные входы-указатели, т.е. вход через ближайшие области применения, и вход через онтологически близкие понятия. Итак, предлагаемый отраслевой переводный тезаурус объединяет преимущества всех лексикографических приемов, максимально учитывая требования пользователей и обеспечивает более удобный и вероятный подход к представлению эквивалентов лексических единиц.

Далее обсуждается, что предлагаемая структура технического переводного тезауруса состоит из трех указателей, или списков: это **пермутационный, иерархический и алфавитный** (лексико-семантический) **указатели**. Основная часть тезауруса является алфавитным указателем. Остальные указатели дополняют его. Пермутационный указатель, отображая вместе все употребляемые в данной области словосочинения одного термина, помогает пользователю выбрать правильное сочетание слов. Кроме того, с помощью пермутационного указателя можно обнаружить более общие или родственные ключевые слова для ключевых словосочетаний, имеющих в тезаурусе. Иерархический указатель представляет парадигматические отношения дескрипторов. Другими словами, иерархический указатель отражает прямое онтологическое, или понятийно-содержательное отношение между терминами. Основные категории иерархического указателя тезауруса включают: (1) **общий признак**; (2) **понятия, связанные с землей и космосом**; (3) **техника**; (4) **материал**; (5) **рабочее условие**; (6) **организм**; (7) **явление**; (8) **процесс**; (9) **свойство**; (10) **экономический фактор**. Однако в прагматическом отношении одно понятие может объединяться со многими другими понятиями. Например, термин *битум* может рассматриваться в связи с техникой его обработки, различными процессами, оборудованием и т.д. Такие прагматические отношения отражаются в алфавитном указателе в соответствующем тематическом разделе. Алфавитный указатель подразделяется на тематические разделы и включает тезаурусные статьи. Каждая статья состоит из ключевого слова (КС), под которым предоставляются все связанные с данной тематикой слова и выражения. Кроме того, иногда даются определения или ссылки на связанные слова или словосочетания. Следовательно, пользователь тезауруса может одновременно рассмотреть данный термин в разных тематических разделах. Иногда также используется описание единицы для большего уточнения её значения. Таким образом, статья тезауруса имеет следующий вид:

**БИТУМ** м. **قير** (I. 1); (II. 2); (IV. 1.1.3), (IV.1.5.)

*Общее название твердых и жидких веществ или продуктов их переработки, состоящих из углеводов, а также из их кислородных, сернистых и азотистых производных.*

*н.в. битуминит*

- с. асфальт آسفالت (IV. 1.1.3), (IV.1.5.)
- в. сырая нефть نفت خام (I. 1); (IV. 1); (V. 1)
- в. твердый углеводород هیدرو کربن جامد (I. 1); (IV. 1.1.3), (IV.1.4)
- в. гудрон ته ماده سنگین (IV. 1.1.3),
- и. природный битум قیر طبیعی (I. 1); (II. 2);
- и. нефтяной битум قیر نفتی (I. 1); (II. 2);
- и. искусственный битум قیر مصنوعی (IV. 1.1.3), (IV.1.5.)
- и. дёготь قیر آسفالتی، قیر مصنوعی (IV. 1.1.3), (IV.1.5.)
- и. гудрон آسفالتوم، ته ماده سنگین، (IV. 1.1.3), (IV.1.4), (IV.1.5.)
- и. смола قیر قطران (IV. 1.1.3), (IV.1.4),
- а. перегонка تقطیر (IV. 1.1.3), (IV.1.4), (IV.1.5.)
- а. крекинг کراکینگ (IV. 1.1.3), (IV.1.4)
- а. очистка масел تصفیه روغن (IV. 1.1.3), (IV.1.2)
- а. вар رزین (IV. 1.1.3),
- гр. битумизация قیر اندود سازی
- гр. битуминизировать قیر اندود کردن
- гр. битумизированное قیر اندود شده
- при. ~ покрытие رویه آسفالتی
- гр. битумный قیر اندود
- при. ~ покрытие اندود قیری

С целью подтверждения принадлежности компонентов к определенному понятию проводился ассоциативный эксперимент, в котором участвовали 19 иранских и русских специалистов нефтегазовой области. В результате они пришли к выводу, что сравнение результатов анализа показывает, что хотя количество информаторов не было достаточным для того, чтобы выполнить всесторонний анализ и прийти к общему, универсальному выводу, но очевидно, что признаки, выделяемые ассоциативным экспериментом для каждой категории или понятия, совпадают в аспекте тематического охвата данного понятия.

В заключении содержатся основные выводы и результаты проведенного исследования, намечаются перспективы составления многоязычных отраслевых тезаурусов.

Приложение к работе наглядно представляет структуру тезаурусной статьи, лексические вхождения и их характеристики в сопоставляемых языках.

По теме диссертации опубликованы следующие работы:

1. **О проблеме перевода слова «обработка» в технической литературе** // Вопросы филологических наук, № 4(27), июль 2007, С. 65-67.
2. **Краткий отчет о XIII мастерской Lexicom и конференции EUGALEX-2008 (Барселона, Испания)** // Вестник, Российского Университета Дружбы Народов; Русский и иностранные языки и методика их преподавания, № 3, 2009, С. 108-111.
3. **Современная лексикографическая деятельность в Иране** // Вестник, Российского Университета Дружбы Народов; Лингвистика, № 1, 2010, С. 90-97.

**Марьям Фаал-Хамеданчи (Иран)**

**Тезаурусное представление терминов нефтегазовой отрасли  
в русском и персидском языках**

В диссертации рассматривается теория и практика лексикографического представления терминологии нефтегазовой отрасли на материале русского и персидского языков. Построение двуязычного русско-персидского/персидско-русского переводного тезауруса на основе информационно-поискового языка впервые разрабатывается для терминологии нефтегазовой отрасли, что особенно важно для нефтедобывающих стран.

Практическое использование результатов исследования приоритетно для лексикографической практики составления дву-/многоязычных отраслевых словарей-тезаурусов. Сопоставительный аспект предполагает возможность использования материалов диссертации в практике преподавания сопоставительной грамматики и лексикологии, сопоставительной типологии, теории перевода, различных спецкурсов по сопоставительной лингвистике и лексикографии.

**Maryam Faal-Hamedanchi (Iran)**

**The Thesaurus Representation of Oil-and-gas Terminology  
in Russian and Persian**

The thesis deals with the theory and practice of lexicographic comparative presentation of oil-and-gas terminology in Russian and Persian. The composition and structuring of bilingual Russian-Persian/Persian-Russian translation thesaurus on the basis of the information query language has been specially elaborated for the oil-and-gas industry which is of vital importance for the oil-producing countries.

Practical usage of the research results is of priority for the lexicographic practice of composing bilingual and multilingual branch thesauruses/ Comparative aspect of the study offers a possibility to use the data and material of the thesis in teaching comparative grammar and lexicology, comparative typology, the theory of translation and a number of optional courses in comparative linguistics and lexicography.

**Подписано в печать: 20.04.2010**

**Заказ № 3587 Тираж - 100 экз.**

**Печать трафаретная.**

**Типография «11-й ФОРМАТ»**

**ИНН 7726330900**

**115230, Москва, Варшавское ш., 36**

**(499) 788-78-56**

**[www.autoreferat.ru](http://www.autoreferat.ru)**

