

ТЕРМИНОЛОГИЯ КАК ОБЪЕКТ ТЕЗАУРУСНОГО МОДЕЛИРОВАНИЯ

С.Е. Быкова

TERMINOLOGY AS OBJECT OF THESAURUS MODELING

S.Y. Bykova

Рассматриваются понятия тезауруса, тезаурусной лексикографии и терминографии. Автор анализирует тезаурусное представление терминологии экологии и экологического права на материале французского языка в электронных терминологических тезаурусах и терминологических банках данных.

Ключевые слова: терминология экологии, терминология экологического права, тезаурус, терминологический банк данных, тезаурусная лексикография, тезаурусная терминография, французский язык.

The article deals with the notions of thesaurus, thesaurus lexicography and terminology. The author of the article analyses thesaurus presentation of terminology of environment and environmental law in on-line French thesaurus and terminological databank.

Keywords: terminology of environment, terminology of environmental law, thesaurus, terminological databank, thesaurus lexicography, thesaurus terminology, the French language.

С развитием информатизации науки и общества возникла необходимость в создании методик формализации языка, что повлекло за собой развитие тезаурусного моделирования. Уже с 70–80-х годов XX века вопросы идеографии вызывали интерес исследователей. Так, к этому периоду относятся труды Р.Ю. Кобрина, В.В. Морковкина, Ю.Н. Караулова и др.

В переводе с греческого термин «тезаурус» («thesauros») означает «сокровище», «сокровищница». В современной лингвистике тезаурусом называют:

1) словарь, в котором максимально полно представлены все слова языка с исчерпывающим перечнем примеров их употребления в текстах;

2) идеографический словарь, в котором показаны семантические отношения (родовидовые, синонимические и др.) между лексическими единицами¹.

Очевидно, что тезаурус в первом значении возможно создать только для мертвых языков. Живые языки, находящиеся в непрерывном развитии, делают задачу создания исчерпывающего тезауруса неосуществимой. Что же касается второго

понимания тезауруса, то именно такие словари все чаще создаются как информационно-поисковые системы.

Лингвистический энциклопедический словарь определяет назначение информационно-поисковых систем следующим образом: «это единообразная замена лексических единиц текста стандартизованными словами и выражениями (дескрипторами) при индексировании документов и использование родовидовых и ассоциативных связей между «дескрипторами» при автоматизированном информационном поиске документов»².

В лексикографии сложилась тенденция относить словари тезаурусного типа к компетенции ее особой отрасли, которая получила название «тезаурусной лексикографии»³. Тезаурусная лексикография занимается решением специфических задач, которые связаны непосредственно с составлением идеографических словарей и информационно-поисковых тезаурусов, с разработкой типологии словарей тезаурусного типа.

Тезаурус является основным инструментом информационно-поисковых систем. Главной ча-

Быкова Светлана Евгеньевна, кандидат филологических наук, доцент, докторант кафедры немецкого и французского языков и методики их преподавания, Нижнетагильская государственная социально-педагогическая академия (г. Нижний Тагил); научный руководитель – В.И. Томашпольский, доктор филологических наук, профессор. E-mail: bykova_svetlana_@mail.ru

Svetlana Y. Bykova, PhD, German and French languages and Methods of Teaching, associate professor, Nizhny Taguil State Socio-pedagogical Academy (Nizhny Taguil); Scientific Supervisor – Valentin. I. Tomashpolsky, PhD, Prof. E-mail: bykova_svetlana_@mail.ru

стью такой системы является словарь терминов, которые упорядочены по алфавитному или алфавитно-гнездовому принципам, указаны их семантические отношения (преимущественно иерархические и ассоциативные, в некоторых случаях – отношения вариантности, синонимии, антонимии). Нередко, особенно в случае терминологического тезауруса, дескрипторы сопровождаются дефиницией, что приближает тезаурус к толковому словарю. Но несомненным отличием тезауруса от традиционного толкового словаря является то, что в тезаурусе лексика упорядочена двумя способами (систематическим и алфавитным), и то, что характерной чертой тезауруса является наличие нескольких входов.

Таким образом, тезаурус расширяет число возможных «точек доступа» к словарному составу языка, увеличивая тем самым размерность семантического пространства лексики: при обязательности алфавитного входа, без которого немислим, очевидно, никакой словарь и который обеспечивает переход от «слова» к «понятию» (или от «знака» к «концепту»), для него конституирующим является вход через «понятие» или «тему», то есть он дает и возможность перехода от «понятия» к «слову» (от «концепта» к «знаку»)⁴.

Практическая ценность тезауруса выражается именно в прагматическом аспекте отношения «концепт» – «знак», в возможности наиболее эффективного поиска слова, выражения или же термина. Также весьма велика значимость тезауруса при обучении иностранному языку в его терминологическом аспекте. Использование в образовательном процессе одно- или многоязычного терминологического тезауруса помогает создать у обучающегося целостную картину понятий изучаемой предметной области.

Кроме информационно-поисковых тезаурусов существуют такие формы хранения информации, как базы данных (БД) и терминологические банки данных (ТБД). БД содержат библиографические или фактографические описания документов различных областей знаний.

ТБД – это электронная оболочка данных, разбитых на узкоспециализированные области и ориентированных на определенные группы пользователей, обладающих различными интересами, знаниями и уровнем подготовки. ТБД сочетает в себе терминологический словарь (одно- или многоязычный), переводной или толковый, и тезаурус. Систематизация данных, содержащихся внутри ТБД, позволяет пользователю получить информацию в наиболее удобной для него форме: алфавитно или системно упорядоченную. Данные, содержащиеся в ТБД, основаны на авторитетных изданиях терминологических словарей или исследовательских работах ведущих специалистов. Часто ТБД являются пополняемыми, что существенно расширяет их возможности по сравнению с терминологическим словарем.

В настоящее время с развитием информационных технологий создается все больше информационно-поисковых тезаурусов и терминологических банков данных по различным отраслям науки. Рассмотрим несколько электронных изданий на французском языке, находящихся в свободном доступе в сети интернет и содержащих терминологию экологии и экологического права. Анализ проводился с точки зрения структуры тезауруса, полноты охвата терминологии, значимости тезауруса для изучения терминологии указанных дисциплин.

*Thésaurus Urbamet*⁵

Библиографическая база данных URBAMET охватывает тематические поля «Урбанизация», «Благоустройство», «Жилище», «Строительство», «Архитектура» и «Оборудование».

Банк данных был создан в 1986 году. Развитие его иерархической структуры стало основой для тезауруса URBAMET. Позднее, в 2002 году, каждый дескриптор был переведен на английский и испанский языки.

Графически тезаурус представляет собой две части окна. Вверху страницы расположена строка поиска. Результат поиска показан в правой части окна. Левая же часть демонстрирует иерархическую структуру тезауруса. Каждый уровень этой структуры – термин и его отношения синонимии и ассоциации. Между уровнями иерархии пользователь может с легкостью перемещаться. При выборе термина щелчком мыши соответствующая словарная статья появляется в правой части окна. Наличие дефиниции отражено в тексте словарной статьи символом, который является гиперссылкой.

Тезаурус регулярно пополняется, а в 2001 году он был основательно переработан. В настоящий момент 4250 дескрипторов входят в состав тезауруса, образуя двадцать четыре семантических категории. Каждый дескриптор относится только к одной из них. Представлено до семи уровней иерархии.

Работа над тезаурусом ведется Центром градостроительной документации Главного управления градостроительства, жилья и строительства Министерства экологии и устойчивого развития при поддержке комиссии тезауруса ассоциации URBAMET.

Термин «*droit de l'environnement*» представлен в тезаурусе в качестве дескриптора. Его место в иерархической структуре тезауруса обозначено так: «*cadre juridique – cadre conceptuel du droit – législation – droit spécifique – droit de l'environnement*». Гипонимов данного термина не представлено, так же как и терминов, связанных с ним ассоциативными отношениями.

Термин «*environnement*» имеет более подробно проработанную структуру. Он входит в семантическую категорию «*environnement – paysage*», имеет гипонимы «*acoustique*», «*patrimoine naturel*», «*politique de l'environnement*», «*protection*», «*qualité de l'environnement*». Ассоциативные отношения не

обозначены. Иерархическая структура каждого из гипонимов представлена довольно подробно. Дефиниции терминов, вошедших в рассмотренные семантические поля, в тезаурусе не представлены.

*Thésaurus des sujets de base du gouvernement du Canada*⁶

Тезаурус базовых понятий правительства Канады является двуязычным (англо-французским). Вошедшая в данный тезаурус терминология отражает темы, представленные в информационных ресурсах правительства Канады. В словник вошли 4859 терминов на французском и 4792 – на английском языке, из которых 2173 являются дескрипторами в каждом из языков. В тезаурусе в разной степени представлены все области знания. Учитывая обширность тематики, для каждой из тем отобраны термины самого общего уровня. Авторы тезауруса сознательно исключили специфическую терминологию узких предметных областей.

Тезаурус прежде всего адресован библиотекарям, специалистам, работающим с документацией в министерствах и ведомствах федерального уровня, и призван облегчить выбор официально признанных терминов для описания тем правительственных ресурсов в сети интернет. Данный тезаурус является нормативным контролируемым источником терминологии.

Французская и английская версии тезауруса разрабатывались одновременно, их статус абсолютно идентичен, то есть ни один из языков не рассматривается как доминирующий или как язык перевода. Семантические отношения в соответствующих группах тезаурусов на французском и английском языках, впрочем, могут отличаться по причине концептуальных и терминологических особенностях этих языков.

Тезаурус включает в себя дескрипторы (предпочтительные термины) и не-дескрипторы (синонимы или нежелательные термины). Дескрипторы связаны между собой иерархическими, ассоциативными отношениями и отношениями межъязыковой эквивалентности. Все термины распределены по крупным семантическим категориям. Некоторые дескрипторы сопровождаются дефиницией.

В тезаурусе представлены девятнадцать категорий: arts, musique, littérature (искусство, музыка, литература), agriculture (сельское хозяйство), économie et industrie (экономика и промышленность), éducation et formation (воспитание и образование), format (формат), gouvernement et vie politique (правительство и политическая жизнь), santé et sécurité (здоровье и безопасность), histoire et archéologie (история и археология), information et communication (информация и коммуникация), travail et emploi (работа и трудоустройство), langue et linguistique (язык и языкознание), droit (право), histoire et science militaire (военная история и наука), nature et environnement (природа и окружающая среда), personnes (персоналии), liens et fonctions (связи и функции), société et culture (об-

щество и культура), sciences et technologie (наука и технологии), transport (транспорт).

Рассматриваемые нами термины экологии и экологического права широко представлены в данном тезаурусе. Семантическая категория «Nature et environnement» включает в себя множество терминов (более 1000), среди которых и «droit de l'environnement», который в качестве дескриптора входит в семантическую категорию «Droit». Каждая словарная статья представляет собой систему гиперссылок, позволяющую просмотреть карточку любого из связанных с заглавным словом терминов. Таким образом, пользователь тезауруса может без труда составить представление о взаимосвязях терминов.

*Le thésaurus juridique belge (UTU)*⁷

Данный тезаурус имеет древовидную структуру и носит универсальный характер, что позволяет упорядочить различные источники права, получить информацию о законодательстве, юриспруденции и теории права. В тезаурус вошли все отрасли права.

Тезаурус общедоступен, он создан для того, чтобы помочь практикующим юристам, издателям, представителям власти классифицировать общепринятым образом различные юридические документы (судебные документы, постановления, законопроекты и т. п.), любую юридическую информацию в базах данных.

Древовидная структура тезауруса отражает избранный его авторами аналитический подход к праву, что отличает данное издание от обычных, алфавитных, списков терминов. При создании структуры тезауруса основой послужили базовые законодательные тексты, которые определили выделение тематических групп и их содержание. Такой подход обеспечил наибольшую объективность предлагаемой информации.

В тезаурусе представлены следующие разделы: droit public (публичное право), droit civil (гражданское право), droit judiciaire (процессуальное право), droit économique (экономическое право), droit pénal (уголовное право), droit fiscal (налоговое право), droit social (социальное право). Внутри каждой отрасли права выделяются группы, расположенные в алфавитном порядке.

Рассматриваемая нами терминология экологии и экологического права вошла в группу Droit public. Представлена двадцать одна подгруппа. Внутри большинства из подгрупп представлены уровни иерархии терминов. Как правило, представлено не более четырех уровней иерархии.

Ввиду специфики данного тезауруса (классификатор юридической документации) подгруппы тезауруса включают в себя разделы, позволяющие разделить документы по их происхождению (международные, европейские, федеральные), а также озаглавленные по названиям единиц административно-территориального деления Бельгии (Flandre, Région wallonne, Région de Bruxelles-capitale).

Фонетика и прикладная лингвистика

В каждой подгруппе базовые законодательные тексты были сгруппированы по ключевым словам в алфавитном порядке. В группе каждого ключевого слова ссылки на законы выстроены в хронологическом порядке.

Данный тезаурус может быть полезен для изучения специфики юридической терминологии Бельгии. К сожалению, ни дефиниции, ни ассоциативные отношения терминов в тезаурусе не представлены.

*Sign@l : Signalement des contenus de périodiques en sciences humaines et sociales*⁸

Тезаурус позволяет осуществить поиск по ключевым словам, по автору, по названию журнала (в базу внесены семьдесят четыре издания с 1983 года, 67475 статей, из которых 14 % доступны он-лайн). Тезаурус создан Лионским Институтом политических наук в 1983 году. Первоначально он носил название FRIPES (каталог библиографической информации по политическим, экономическим и общественным наукам) и разрабатывался в партнерстве с региональными университетами и организациями. Благодаря деятельности Института политических наук каталог со временем получил широкое распространение в сети интернет.

В 1998 году Институтом была создана система SOMREV, представляющая собой базу данных содержаний журналов. В 2011 году две базы, SOMREV и FRIPES были объединены. Так появился тезаурус «Sign@l».

Поиск информации в тезаурусе можно осуществить путем ввода с клавиатуры ключевых слов в строке поиска или при помощи иерархического списка терминов. Этот список включает в себя четыре тематические группы: ANTHROPO (персоналии), GEO (топонимы), MATIERE (предметные области), ORGA (организации).

Рассматриваемые в нашей работе термины экологии и экологического права включены в группу MATIERE, подгруппу «Environnement – Protection de l'environnement – Ressources naturelles – Agriculture – Énergie» (окружающая среда – защита окружающей среды – природные ресурсы – сельское хозяйство – энергетика). Как видно, подгруппа охватывает довольно широкий круг тем. Тем не менее в ней представлено сорок четыре термина, относящиеся к интересующим нас семантическим группам. В тезаурусе представлены только иерархические отношения, при этом они обнаруживаются только внутри микрогрупп. Так, подгруппы указанной выше группы MATIERE расположены в алфавитном порядке, так же как микрогруппы подгруппы Environnement – Protection de l'environnement – Ressources naturelles – Agriculture – Énergie. Но внутри микрогрупп наблюдается иерархия терминов. Ассоциативные отношения не представлены в тезаурусе совсем. Однако ценность данного тезауруса довольно велика, так как он позволяет получить доступ к значительному числу статей по искомой тематике, причем многие из

этих статей представлены в полнотекстовом варианте, остальные же – в виде полной библиографической справки.

*Ressources documentaires des CAUE (Conseils d'Architecture, d'Urbanisme et de l'Environnement)*⁹

Тезаурус Совета по архитектуре, градостроительству и защите окружающей среды. Представляет собой систему доступа к документам данной организации. Включает в себя двадцать пять семантических полей: сельское хозяйство, благоустройство, архитектура, искусство, энергетика, строительство, право, вода, экология, экономика, образование, окружающая среда, оборудование, лес, физическая география, жилище, методы исследования, служебные слова, пейзаж, вредное воздействие, здравоохранение, гуманитарные науки, туризм и отдых, транспорт, урбанизация.

В тезаурусе представлены только иерархические отношения терминов (до пяти уровней иерархии). В семантическом поле термина Environnement рассматриваются вопросы экологического права (тридцать девять терминов), устойчивого развития (пять терминов), отношений «человек – природа», природного достояния, планирования в сфере охраны окружающей среды, качества окружающей среды, теории окружающей среды, экологической политики, политической экологии, экономики окружающей среды.

Существенным недостатком тезауруса является отсутствие информации об ассоциативных связях терминов, отношениях синонимии, а также отсутствие дефиниций терминов. Число отобранных терминов представляется недостаточным для полного охвата заявленных тем.

*Thésaurus de Statistique Canada*¹⁰

Статистический тезаурус Канады был создан с целью доступа к информационным ресурсам через тематический поиск. Тезаурус – это инструмент индексации и документационного поиска на двух языках, с четкой структурой, разработанный согласно международным стандартам. Он состоит из дескрипторов и не-дескрипторов, отражает их иерархические, ассоциативные и нормативные отношения.

Содержание данного тезауруса постоянно пополняется и отражает темы статистических ресурсов Канады. Термины отобраны из различных документов, включая публикации и базы данных. В тезаурус не вошли географические названия, имена собственные, аббревиатуры (кроме общепринятых).

В тезаурусе равнозначны французский и английский языки. Французская и английская версии тезауруса существуют параллельно, соответствия проводятся в обязательном порядке только для дескрипторов. Во французской части тезауруса дескрипторы представлены в форме мужского рода (например, «acteurs», а не «actrices»).

Как правило, термины – это существительные или словосочетания с ними, в единственном числе,

с прямым порядком слов (в случае словосочетания). В случае необходимости, приводится заметка с краткой дефиницией термина и рекомендацией по его правильному употреблению.

Как уже было упомянуто выше, в тезаурусе представлены иерархические (родовые и видовые термины), ассоциативные (связанные термины) и нормативные («употребляется для...») отношения. Родовым термином является дескриптор, которому подчинен один или несколько других дескрипторов. Видовой термин – дескриптор, подчиненный одному или нескольким другим дескрипторам. Связанный ассоциативными отношениями термин – это дескриптор, не входящий в иерархические отношения с данным дескриптором тезауруса. Заметка об употреблении термина отражает его эквивалентность, отношения синонимии, неполной синонимии и лексической вариантности.

Доступ в тезаурус может осуществляться тремя способами: поиском по слову или словосочетанию (нужно напечатать его в окне поиска), при помощи алфавитного списка и при помощи тематического списка. В тезаурус вошли тридцать две тематические группы: *peuples autochtones* (коренные народы); *agriculture* (сельское хозяйство); *commerce, consommateurs et services de propriété* (торговля, потребитель, сфера услуг); *rendement des entreprises et appartenance* (доходность предприятий и их принадлежность); *enfants et jeunes* (дети и молодежь); *construction* (строительство); *crime et justice* (преступление и правосудие); *culture et loisirs* (культура и отдых); *comptes économiques* (экономическая отчетность); *éducation, formation et apprentissage* (общее и профессиональное образование); *énergie* (энергия); *environnement* (окружающая среда); *diversité ethnique et immigration* (этническое разнообразие и иммиграция); *familles, ménages et logement* (семья, домашнее хозяйство и жилище); *gouvernement* (правительство); *santé* (здоровье); *revenu, pensions, dépenses et richesse* (доходы, пенсии, расходы, богатство); *technologies de l'information et des communications* (информационные и коммуникационные технологии); *commerce international* (международная торговля); *travail* (труд); *langues* (языки); *fabrication* (производство); *population et démographie* (население и демография); *prix et indices des prix* (цены и индекс цен); *référence*

(ссылки); *commerce de détail et de gros* (розничная и оптовая торговля); *science et technologie* (наука и технология); *aînés* (старшее поколение); *société et communauté* (общество); *méthodes statistiques* (статистические методы); *transport* (транспорт); *voyages et tourisme* (путешествия и туризм).

В рамках изучаемой нами темы рассмотрим тематическую группу «*Environnement*». Она состоит из шести подгрупп: общие понятия, воздух и климат, защита окружающей среды, качество окружающей среды, природные ресурсы, загрязнение и отходы. В группу «Общие понятия» входит пятнадцать ключевых терминов, таких как «*écologie*» (экология), «*environnement*» (окружающая среда), «*législation sur l'environnement*» (законодательство в сфере охраны окружающей среды) и др. Еще 137 терминов вошли в состав остальных пяти групп.

В данном тезаурусе термины не сопровождаются толкованием, так как само назначение данного тезауруса (доступ к базам публикаций) этого не предполагает. Сами публикации доступны для просмотра на сайте или для скачивания в формате Pdf.

Рассмотренные выше тезаурусы и ТБД подтверждают тот факт, что тезаурусное моделирование терминологии имеет практическое значение как в сфере ее упорядочения, так и в организации работы с библиографическими данными, текстами нормативных и других документов. Создание терминологических тезаурусов является перспективным направлением современной терминографии.

¹ Лингвистический энциклопедический словарь / под ред. В.Н. Ярцевой. М.: Советская энциклопедия, 1990. С. 506.

² Там же. С. 506–507.

³ Герд, А.С. Научно-техническая лексикография / А.С. Герд // Прикладное языкознание : учеб. СПб.: Изд-во С.-Петерб. гос. ун-та, 1996. С. 171.

⁴ Горбунов, Ю. И. Тезаурусное моделирование французской лингвистической терминологии: дис. ... д-ра филол. наук / Ю.И. Горбунов. Тольятти, 2004. С. 61.

⁵ URL: www.urbamet.com

⁶ URL: www.thesaurus.gc.ca

⁷ URL: www.thesaurusuniversel.be

⁸ URL: <http://doc.sciencespo-lyon.fr/Signal/>

⁹ URL: <http://www.archi.fr>

¹⁰ URL: http://www47.statcan.ca/th_r000_f.htm

Поступила в редакцию 3 октября 2011 г.