

## ТЕОРИЯ И ПРАКТИКА ПЕРЕВОДА; НОВЕЙШИЕ КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ В ФИЛОЛОГИИ

УДК 81'33

### СОСТАВЛЕНИЕ ТЕЗАУРУСА ПО ЭКОНОМИКЕ НА БАЗЕ КОРПУСА УЧЕБНЫХ ТЕКСТОВ ПО СПЕЦИАЛЬНОСТИ

*Кривонос Е. А.*

*Белорусский государственный университет, г. Минск, Беларусь  
E-mail: kerris@inbox.ru*

Статья посвящена представлению созданного «Тезауруса по экономике для иностранных граждан, изучающих русский язык (этап довузовской подготовки)» (ТЭИ-0), а также описанию этапов работы над словарем. ТЭИ-0 был разработан на базе исследовательского корпуса текстов по экономике, в котором содержится 41 учебный текст. Также в статье обоснована актуальность создания тезаурусного представления предметных областей.

**Ключевые слова:** тезаурус, корпус текстов, экономическая терминология

**Постановка проблемы.** В связи с увеличением информационного потока ежедневно возникает необходимость совершенствования как систем обработки и хранения информации, так и навигации в этих системах. Одним из путей решения проблемы в данном случае является систематизация неструктурированных данных. Представление знаний в формализованном виде не только значительно ускоряет процессы работы с информацией, но и повышает их продуктивность.

В свете указанного направления развития информационных технологий работа по моделированию предметных областей и представлению полученных данных в виде тезаурусов является весьма востребованным видом деятельности. Результат представляет собой гипертекст, который отражает иерархически организованную семантическую структуру определенной предметной области.

Представление тезаурусного описания предметных областей гуманитарных наук является более сложной задачей для реализации, чем естественно-научных дисциплин, в силу различных причин: наличия не единственной классификации объектов, недоопределенности некоторых объектов. Все это порождает терминологический разнобой, последствия которого проявляются в первую очередь при прикладном использовании данной информации, например в системах автоматического перевода, в сфере компьютерной лексикографии, компьютерной лингводидактики в связи с более широким распространением дистанционных форм обучения (e-learning), охватывающих с каждым все большее количество людей. Процесс глоба-

лизации обучения, когда стираются границы между странами, особенно в Интернет-пространстве, требует, насколько это возможно, однозначного и выверенного подхода к представлению о конкретной предметной области.

Актуальность данной темы подтверждается многочисленными исследованиями по системному описанию предметных областей, например: «Функциональная модель двуязычного экологического словаря-тезауруса» М. А. Ковязиной, «Лингвистическая модель двуязычного электронного текстоориентированного словаря юридических терминов : на материале Конституции Российской Федерации, Конституции Французской Республики и Конвенции о защите Прав человека и основных свобод» А.В. Русаковой, «Тезаурусное моделирование английских грамматических терминов морфологии и морфосинтаксиса» Е.Ю. Горбунова. В книге «Семантика информационных технологий: Опыты словарно-тезаурусного описания» Ю.Н. Филипповича и А.В. Прохорова [1] описана методика и алгоритмы дистрибутивно-статистических методов анализа естественно-языковых описаний предметных областей и построения на их основе словарей-тезаурусов, подробно рассмотрен программный комплекс Интерлекс, использованный для исследования и приведен семантический словарь-тезаурус предметной области «Информационные технологии».

В данный момент тезаурусному описанию русскоязычной экономической терминологии посвящено не так много исследований, несмотря на то что экономика является одной из основ жизнедеятельности общества. Это работы З.И. Комаровой и А.А. Прошиной «Моделирование двуязычного словаря-тезауруса по экономике» [2], Ю.В. Рождественского «Общество. Семиотика. Экономика. Культура. Образование: словарь терминов (Общеобразовательный тезаурус)» [3], Информационно-поисковый тезаурус ИНИОН РАН по экономике и демографии [4].

Монография «Моделирование двуязычного словаря-тезауруса по экономике» [2] посвящена описанию первого опыта сопоставительного моделирования экономической терминотерминологии в русском и немецком языках с целью терминографирования. Особенность предлагаемой модели заключается в ее универсальности, что позволяет совместить функции толкового, переводного и идеографического терминологического словаря.

Одной из главных причин небольшого количества работ, посвященных моделированию экономической концептосферы, следует назвать состояние самой русской экономической терминологии, которая в связи с политическими событиями конца 20 – начала 21 века подверглась глобальной трансформации и в данный период еще находится в стадии становления. Все вышеперечисленные исследования были ориентированы на максимально полное описание русской экономической терминологии. Но мы бы хотели обратить внимание на необходимость создания тезаурусного описания концептосферы экономики в несколько ином аспекте. По данным исследования рынка образовательных услуг в ближайшие пять лет прогнозируется существенное его увеличение. В первую очередь прирост рынка ожидается за счет иностранных граждан, которые пожелают получить высшее образование на русском языке. С учетом развития современного общества повышается и уровень требований к учебно-методическому обеспечению образовательного процесса. Поэтому удовлетворить все возрастающие потребности без использования методов автома-

тической обработки текста в ближайшее время уже будет невозможно. Цель активного внедрения новых информационных технологий автоматического анализа текста на этапе подготовки учебно-методического материала – получение новых интеллектуальных продуктов. Одним из таких продуктов является ряд специализированных адаптированных тезаурусов, которые будут соответствовать требованиям, предъявляемым к уровням владения русским языком. Отсутствие четко описанных понятийных минимумов по специальностям порождает такие проблемы, как-то: субъективность и интуитивность в определении языковой компетенции иностранного специалиста при тестировании по языку специальности, отсутствие критериев определения соответствия текста по специальности уровню владения языком и т.д. Создание же тезаурусов, ориентированных именно на иностранных граждан, позволит начать решать указанные проблемы, а также не только окажет неоценимую помощь в методическом обеспечении учебного процесса для иноязычной аудитории на более высоком профессиональном уровне, но и будет способствовать усвоению иностранцем корректной семантической структуры понятийной области, исключая факты интуитивности или непонимания при восприятии информации на неродном языке. Следует отметить, что наиболее важны подобные тезаурусные описания на этапе получения первичных знаний о системе любой понятийной области. Поэтому представляется весьма актуальным в первую очередь проведение работы по созданию тезаурусов, которые ориентированы на иностранных граждан, обучающихся на подготовительных отделениях вузов. Подобные тезаурусы должны быть созданы для различных специальностей.

Целью нашей статьи является представление созданного «Тезауруса по экономике для иностранных граждан, изучающих русский язык (этап довузовской подготовки)» (ТЭИ-0), а также описание этапов работы над словарем.

Материалом для исследования явились учебные тексты по экономике на русском языке из учебно-методического комплекса «Экономика» для иностранных граждан [5], активно применяемого в учебном процессе на факультете доуниверситетского образования БГУ. Всего в исследовательский корпус текстов по экономике для этапа довузовского обучения (ИКТ-0) был включен 41 учебный текст.

Работа по разработке ТЭИ-0 была разбита на несколько этапов. На первом этапе работы над ТЭИ-0 тексты из ИКТ-0 были обработаны. В результате содержание каждого текста было представлено в виде семантической сети, и для каждого текста создан минитезаурус. На втором этапе полученная в ходе обработки текстов из ИКТ-0 информация была обобщена, что позволило определить круг терминологических единиц, подлежащих упорядочиванию. Следующим этапом в работе явилось структурирование информации в соответствии с моделью тезауруса, описание которой представлено ниже. Результатом деятельности является ТЭИ-0 – «Тезаурус по экономике для иностранных граждан, изучающих русский язык (этап довузовской подготовки)».

Опытная модель построения двуязычного словаря-тезауруса по экономике, предназначенного в первую очередь для целевого использования в учебном процессе, подробно описана в монографии З.И. Комаровой и А.А. Прошиной [2]. Авторы отмечают, что моделируемый словарь сочетает в себе типологические характеристики идеографического, толкового и переводного словаря и выполняет все основ-

ные словарные функции: систематизирующую, справочную, нормативную и учебную. Наличие в словаре дефиниций необходимо, потому что экономическая терминосфера находится в стадии формирования, что влечет за собой некоторую неупорядоченность. Кроме того, созданный нами ТЭИ-0 ориентирован на иностранных студентов, поэтому в учебных целях наличие полной дефиниции, готовой для воспроизведения мы считаем необходимым. В круг наших задач не входит создание двуязычного тезауруса, поэтому мы заимствуем предложенную авторами модель, за исключением компонентов, связанных с переводом на иностранный язык.

Модель тезауруса включает в себя следующие части: иерархический систематический указатель; алфавитный лексико-семантический указатель; пермутационный указатель; алфавитный указатель.

Всего в ТЭИ-0 содержится 328 элементов (терминов и терминологических словосочетаний).

В **иерархическом систематическом указателе** представлены только термины и терминологические словосочетания, которые расположены в соответствии с входением в понятийное поле.

В ТЭИ-0 концептосфера *Экономика* делится на две большие концептуальные области *Экономические науки* и *Хозяйственная деятельность людей*. Понятие *экономика* имеет числовой код 0, в силу того что является точкой отсчета для осуществления кодирования других единиц с целью отражения положения элемента в терминологической системе. В соответствии с нашими потребностями мы используем для обозначения концептуальных областей буквенный код – латинские буквы: для области *Экономические науки* – код *N*, а для области *Хозяйственная деятельность* – код *D*. Третий и последующие уровни иерархии в ТЭИ-0 реализуются только для области *N*. В силу того что область *N* предполагает наличие комплекса наук, то, на наш взгляд, для каждой уместно использовать не числовой код, а буквенный. Вследствие данного обстоятельства концептуальные области третьего уровня иерархии рациональнее кодировать также с помощью латинских букв, исключая *N* и *D*, закрепленные за вторым уровнем иерархии. Учебной программой ФДО предусмотрено изучение основ только одной из комплекса экономических наук – экономической теории; ей присвоен код *R*. На четвертом уровне мы выделили области *R 1* *Общая экономическая теория*, *R 2* *Микроэкономика*, *R 3* *Макроэкономика*, *R 4* *Мировая экономика*. При распределении отобранных терминов и терминологических словосочетаний мы создали структуру, в которой 12 уровней. Приведем в качестве примера фрагмент иерархического указателя:

- R 1.6 Производство
- R 1.6.1 Факторы производства
- R 1.6.1.1 Материальные ресурсы
- R 1.6.1.1.1 Земля
- R 1.6.1.1.2 Капитал
- R 1.6.1.1.2.1 Вещественный капитал
- R 1.6.1.1.2.1.1 Средства производства
- R 1.6.1.1.2.1.1.1 Средства труда
- R 1.6.1.1.2.1.1.2 Предметы труда

- R 1.6.1.1.2.1.1.2.1 Сырье
- R 1.6.1.1.2.2 Человеческий капитал
- R 1.6.1.1.2.3оборотный капитал
- R 1.6.1.1.2.4 Основной капитал
- R 1.6.1.1.2.5 Денежный капитал
- R 1.6.1.2 Людские ресурсы
- R 1.6.1.2.1 Труд
- R 1.6.1.2.1.1 Рабочая сила
- R 1.6.1.2.2 Предпринимательская деятельность

Мы бы хотели уточнить, что нашей задачей является не максимально полное описание конкретной концептуальной области, а определение и представление минимума специальной терминологической лексики, необходимой для понимания адаптированных текстов по экономике. И на включение терминологической единицы в минимум только должен влиять факт нахождения ее в тексте, включенном в ИКТ-0.

В алфавитном лексико-семантическом указателе содержатся полные словарные статьи, упорядоченные в алфавитном порядке. Все статьи строятся по образцу: 1. дескриптор, выделенный жирным шрифтом; 2. дефиниция дескриптора, выделенная курсивом; 3. вышестоящий дескриптор (дескрипторы) с пометой **BT**; 4. нижестоящий дескриптор (дескрипторы) с пометой **NT**; 5. синонимы (помета **S**); 6. термины, обозначающие противоположные понятия (помета **A**); 7. буквенный, числовой или буквенно-числовой код, указывающий на положение терминоединицы в системе. Приведем примеры словарных статей:

**Инфляция** – это обесценивание денег, которые находятся в обращении; падение их покупательной способности

**BT** макроэкономические проблемы

**NT** типы инфляции

**A** дефляция

**R 3.6.2**

**Национализация** – превращение объекта, имущественного комплекса, экономического ресурса или предприятия из собственности частной в собственность государства

**BT** преобразование форм собственности

**R 1.9.6.1**

**Предельные издержки производства** – издержки, связанные с производством еще одной единицы продукции

**BT** виды издержек производства

**S** дополнительные издержки производства

**R 2.5.2.2.5**

Подобное представление позволяет не только актуализировать иерархические отношения между элементами терминосистемы, но эксплицировать отношения синонимии и антонимии, что в комплексе оптимизирует процесс восприятия новой информации: обучаемый не затрачивает время на установление связей между понятиями – они уже установлены и наглядно продемонстрированы.

Кроме того, лексико-семантический указатель дает возможность оценить, насколько развиты отношения синонимии и антонимии понятий.

**Пермутационный указатель** содержит только термины, упорядоченные по алфавиту для каждого слова, в составе термина, т.е. формируются словообразовательные гнезда. Поэтому терминологические сочетания могут дублироваться в нескольких гнездах. Подобные пермутационные указатели облегчают поиск многословных терминов.

#### **Макроэкономика**

Показатели **макроэкономики**

**Макроэкономические проблемы**

**Макроэкономическое равновесие**

Идеальное **макроэкономическое равновесие**

Общее **макроэкономическое равновесие**

Реальное **макроэкономическое равновесие**

Частичное **макроэкономическое равновесие**

Кроме того, собранный и сгруппированный подобным образом материал является источником для описания структуры терминологических словосочетаний.

В **алфавитном указателе** содержатся только термины, расположенные в алфавитном порядке. Для каждого термина указан буквенно-числовой код.

**Выводы и перспектива.** Исходя из вышеизложенного, мы можем сделать следующие выводы. Во-первых, содержание и организация частей созданного ТЭИ-0 в полном объеме позволяют реализовать заявленные систематизирующую, справочную, нормативную и учебную функции. ТЭИ-0 необходим как для облегчения процесса усвоения основных понятий экономической терминологии на русском языке, так и для осознания связей и отношений между элементами терминологической системы. Во-вторых, ТЭИ-0 – является лексическим минимумом по языку специальности (экономический профиль) для этапа довузовского обучения иностранных слушателей.

Следует отметить, что в данный момент ощущается острая необходимость в создании подобных тезаурусов по различным специальностям с учетом уровней владения русским языком. В первую очередь важно удовлетворить запросы на данные минимумы для этапа довузовского обучения иностранных граждан, в силу того что именно отсутствие хороших основ знаний в начале обучения является тем негативным фактором, которое тормозит дальнейшее обучение любого индивидуума.

#### **Список литературы**

1. Филиппович, Ю.Н., Прохоров, А.В. Семантика информационных технологий: Опыт словарно-тезаурусного описания. С предисловием А.И. Новикова / Ю.Н. Филиппович, А.В. Прохоров. – М.: МГУП, 2002. – 368 с.

2. Комарова, З.И., Прошина, А.А. Моделирование двуязычного словаря-тезауруса по экономике : Монография / З.И. Комарова, А.А. Прошина. – Екатеринбург: Форум-книга, 2009. – 276 с.
3. Рождественский, Ю.В. Общество. Семиотика. Экономика. Культура. Образование: словарь терминов (Общеобразовательный тезаурус) / Ю.В. Рождественский. – М.: Флинта : Наука, 2002. – 108 с.
4. Информационно-поисковый тезаурус ИНИОН по экономике и демографии / РАН. ИНИОН. Центр информатизации. Отдел научно-библиографической информации; Архангельская В.А. при участии Мдивани Р.Р. (классификационный указатель). – М.: ИНИОН РАН, 2001. – 494 с.
5. Родионов, В.Н. Экономика : учеб.-метод. комплекс для иностранных слушателей фак. доунив. образования БГУ / В.Н. Родионов, А.В. Воробьев, Т.З. Сафронова. – Мн.: БГУ, 2006. – 240 с.

**Кривонос Е. А. Складання тезауруса по економіці на базі корпусу учбових текстів по спеціальності / Е. А. Кривонос // Вчені записки Таврійського національного університету ім. В. І. Вернадського. Серія «Філологія. Соціальні комунікації». – 2010. – Т. 23 (62), № 3. – С. 196-202.**

Стаття присвячена представленню створеного «Тезауруса по економіці для іноземних громадян, які вивчають російську мову (етап довузовської підготовки)», а також описанню етапів роботи з словником. Даний Тезаурус був розроблений на базі дослідного корпусу текстів по економіці, якій вміщують 41 учбовий текст. Також в статті обоснована актуальність тезаурусного представлення предметних галузей.

*Ключові слова:* тезаурус, корпус текстів, термінологія по економіці.

**Krivoson E. A. Drafting of thesaurus on economy on base of corps of educational texts on speciality / E. A. Krivoson // Scientific Notes of Taurida V. I. Vernadsky National University. – Series: Philology. Social communications. – 2010. – Vol. 23 (62), No 3. – P. 196-202.**

The article is about the presentation of the created «Thesaurus on economy for foreign students, studying Russian, being of the pre-education stage», and also about the description of the stages of work on the dictionary. This Thesaurus has been developed on the basis of research into the corps of texts on economy, which included the 41 educational text. Also the actuality of Tezaurus presentation of any subject domains has been demonstrated.

*Keywords:* thesaurus, corps of texts, economic terminology

*Поступила в редакцію 01.09.2010 г.*