Л. Н. Пирумова, заместитель директора, к. пед. н.
Федеральное государственное бюджетное научное учреждение
«Центральная научная сельскохозяйственная библиотека», г. Москва
pln@cnsheb.ru

УДК 025.4:63

10.31676/2073-4948-2018-52-181-187

НОРМАЛИЗАЦИЯ И ГАРМОНИЗАЦИЯ ОТРАСЛЕВОЙ ТЕРМИНОЛОГИИ В ОТРАСЛЕВОМ ТЕЗАУРУСЕ: ЦЕЛИ И ЗАДАЧИ

Резюме. Работа по обогащению контента тезауруса новой отраслевой лексикой направлена на его актуализацию в соответствии с современным состоянием науки и практики. Тезаурус по сельскому хозяйству и продовольствию, создаваемый в ФГБНУ ЦНСХБ, включает 49659 терминов по всем отраслям АПК и отвечает требованиям ГОСТа. В число его функций входят в том числе: контроль и нормализация лексики по сельскому хозяйству и продовольствию и функция терминологического справочного пособия в области сельского хозяйства и продовольствия. Рассматриваются научные исследования и процессы, проводимые по нормализации научной лексики, включаемой, тезаурус: отбор терминов, лексикографическая обработка, систематизация и группировка лексики, создание иерархических деревьев и классификационных схем.

Ключевые слова: терминология, научная лексика, стандартизация, тезаурус, АПК, ЦНСХБ.

Summary. Enrichment of thesaurus content with new professional vocabulary is oriented toward its updating according to the current state of science and practice. The agriculture and food thesaurus created in the Central Scientific Agricultural Library of 49659 terms in all branches of Agricultural Industrial Complex and meets the requirements of State Standard. Among its functions are: control and normalization of agriculture and food vocabulary, and functioning as the terminological handbook in the field of agriculture and food. The scientific research and processes carried out for normalization of the scientific vocabulary included into thesaurus are shown: term selection, lexicographic processing, vocabulary systematization and grouping, creation of hierarchical trees and classification schemes.

Keywords: terminology, scientific vocabulary, standardization, thesaurus, Agricultural Industrial Complex, Central Scientific Agricultural Library.

Введение

Язык науки – это система понятий, знаков, символов, создаваемая и используемая в науке для получения, обработки, хранения и применения знаний. Язык науки стремится к точности и однозначности выражений. Основой языка конкретных наук является естественный язык, дополненный специальными терминами, знаками и символами. Специфика отдельных отраслей науки выражается терминами, которые в форме слова, словосочетания, аббревиатуры, символа, сочетания слова и символа, являются определением отдельного понятия научной или профессиональной

деятельности, отражая наиболее существенные его признаки [1]. При этом понятие, обозначаемое термином, изменяется вместе с изменениями представлений о предметной области, поэтому с течением времени может меняться формулировка термина, раскрывающая содержание понятия. Кроме того, одно и то же понятие можно описать разными словами одного и того же естественного языка, присвоить ему разные термины и тогда возможно непонимание и ошибки в восприятии, отображении событий, действий, предметов и т.д. Еще большее непонимание возникает, когда ученые говорят на разных естественных языках: в научных статьях ученых разных стран достаточно часто одни и те же понятия выражаются разными терминами.

Поэтому так важно упорядочение научной терминологии, ее унификация и гармонизация, приведение в систему – терминосистему с целью обеспечения взаимопонимания ученых и упрощения научного обмена знаниями, достижениями, информацией на национальном и международном уровнях. Целью нормализации и гармонизации является создание единого языка науки для обмена и распространения знаний. Научная терминология собирается в терминологических словарях и справочниках. Наиболее сложным среди них является тезаурус [2], который является не просто терминологическим справочником, но и терминосистемой, т.е. системой научных понятий с их связями. Системой, в которой термины – лексические единицы, обогащены связями синонимии и приведены в структуру, не противоречащую системам научных классификаций (наоборот, опираются и базируются на них) и обеспечивающую возможность построения гибких стратегий информационного поиска. Потому, что кроме того, что тезаурус – это терминологический справочник, он в то же время и искусственный, специально разработанный для поиска в автоматизированных системах, информационно-поисковый язык [3]. Тезаурус можно рассматривать также как информационно-поисковую систему, в которой каждому слову соответствует определенный дескриптор, представляющий собой слово или символ, по которому можно осуществить поиск информации.

Результаты исследований

Информационно-поисковый тезаурус по сельскому хозяйству и продовольствию (ИПТ) создан и постоянно пополняется и актуализируется в ЦНСХБ в соответствии с ГОСТ [4]. ИПТ включает терминологию по всем отраслям АПК. Объем ИПТ 49659 лексических единиц (ЛЕ). Более 18200 ЛЕ являются научными (латынь) наименованиями организмов. Отрасли знаний представлены в тезаурусе следующим образом: ветеринария – 26%; пищевая промышленность – 17%; защита растений – 19%, животноводство – 10%; лесное хозяйство – 8%; садоводство и виноградарство – 5%; генетика и селекция – 6%; декоративные культуры – 3%; техническое обеспечение АПК – 4%; экономика сельского хозяйства – 2%.

Таким образом, по садоводству, виноградарству, декоративным культурам в тезаурусе около 8% терминов (3962).

При создании и обогащении контента ИПТ новой лексикой учитывались требования, которые отвечают задаче обеспечения полноты и точности поиска:

Однозначность – каждая запись должна иметь только один смысл (т.е. в тезаурусе устраняются такие недостатки, с точки зрения поиска, естественно-го языка как полисемия и омонимия, синонимия).

Явное выражение полезных для поиска семантических отношений между словами (логических отношений и психологических ассоциаций, что делает ИПТ справочником, позволяющим ориентироваться в области знания, устанавливать виды связей и взаимодействия между отдельными понятиями).

Возможность корректировки и дополнения ИПТ, т.е. гибкость и способность включать новые понятия без принципиального изменения его структуры.

Удобство пользования

Способность точно идентифицировать предмет, отличить его особенности и описать его с необходимой степенью детализации и глубины. Это зависит от лексического состава, от развития системы синтаксических отношений между ЛЕ. Все эти качества необходимые для обеспечения релевантного тематического поиска, важны и для терминологического справочника.

Семантическая сила, богатство ИПТ выражается терминологическим запасом, дифференцированным характером парадигматических отношений, развитостью структуры. Лексические единицы в ИПТ – слова на естественном языке, устойчивые словосочетания, аббревиатуры, символы, даты, общепринятые сокращения, лексически значимые компоненты сложных слов.

Обогащение контента ИПТ новой лексикой проводилось таким образом, чтобы не нарушались и не искажались его функции:

- обеспечение индексирования документов и запросов;
- отражение парадигматических отношений (отношения общности и противопоставления значений и использования), существующих между ЛЕ;
- контроль и нормализация лексики по сельскому хозяйству и продовольствию;
- обеспечение единого и формализованного представления информации в информационно-поисковых системах;
- функция терминологического справочного пособия в области сельского хозяйства и продовольствия;
- формально-логический контроль терминов индексирования;
- автоматизированное расширение ПОД.

Работа над обогащением контента ИПТ новой лексикой включает несколько этапов, которые проходят все термины, включаемые в него:

1. Отбор терминов путем просмотра, изучения, анализа, сопоставления научных документов из входного потока, который происходит по следующим критериям:

- частота появления в индексируемых документах и поисковых запросах;
- полезность для поиска информации;
- наличие в авторитетных справочниках, терминологических стандартах;
- наличие в тезаурусах зарубежных и международных систем по сельскому хозяйству и продовольствию.

2. Процесс нормализации лексики (лексикографическая обработка)- выбор наиболее удобных и точных вариантов употребления терминов и утверждение их в качестве предпочтительных, выбор оптимальной формы терминов. При этом происходит унификация терминов, которая призвана обеспечить однозначное соответствие между системой понятий и терминосистемой: одному понятию в терминосистеме ИПТ должен соответствовать только один термин. Терминосистема должна наглядно отражать систему научных понятий с их связями. Поэтому основная задача оптимизации терминологии - поиск оптимальной формы термина, которая непосредственно или косвенно отражает основные классификационные признаки называемого им понятия. С этой же целью установлены унифицированные формы записи для ИПТ: род, падеж, число, разрешение использования словосочетаний и целостных понятий, определения порядка записей в них (наличие инверсии). Каждый термин проходит научную обработку, включающую: выявление синонимии, определение статуса в ИПТ, для чего выясняются его происхождение, история использования, страна происхождения, эволюция формулировки и т.д. Для выражения статуса используются различные уточняющие или ограничительные пометки. Для преодоления языковой неоднозначности (полисемия, омонимия) в тезаурусе используются реляторы (краткое уточнение термина, заключенное в круглых скобках). Для этой же цели используются разные формы грамматического числа и лексические примечания методического характера. Лексические примечания – это свободный текст, размещенный в словарной статье для уточнения понятия, отражаемого дескриптором, в целях облегчения выбора термина. Таким образом, нормализация имеет две стороны: унификацию, направленную на упорядочение содержания терминов, и оптимизацию, имеющую целью выбор оптимальной формы терминов.

3. Систематизация и группировка лексики. ИПТ сложная терминологическая система, между элементами которой – лексическими единицами – существуют различные виды связи – отношения. Смысловые (парадигматические отношения) между дескрипторами (разрешенными к использованию терминами) и аскрипторами (запрещенными к использованию терминами) определяют структуру тезауруса. Основными

методологическими принципами формирования парадигматической структуры ИПТ являются: категоризация лексического состава; построение классификационных схем основных понятий.

Построение классификационных схем (систематизация понятий данной области знания по категориям и построение классификационных схем понятий, в результате чего становятся наглядными существенные признаки понятий; уточнение на основании классификационных схем существующих дефиниций (научных определений) понятий или создание новых дефиниций). Категоризация лексики – это разбиение лексических единиц на определенные тематические группы, представленные общими терминами этих групп. Важное условие категоризации – непересекаемость понятий, отнесенных к разным категориям.

Построение классификационных схем – это установление парадигматических, т.е. внеконтекстных логических связей для ЛЕ. Результатом построения классификационной схемы является словарная статья данной ЛЕ, она может содержать: лексическое примечание; аскрипторы; вышестоящие термины; нижестоящие термины; ассоциативные термины [5]. Между ЛЕ устанавливаются иерархические отношения (отношения подчинения), объем одного из которых составляет часть другого (отношения типа «род-вид», «целое-часть»). Для основных понятий выделяются наиболее существенные их признаки, которые выступают в качестве основания деления широких понятий на более узкие, которые также делятся на более узкие. Так строится классификационная схема понятия с иерархической структурой, так называемое иерархическое дерево. Иерархические деревья, отображают родо-видовые отношения между дескрипторами. ЛЕ имеют отношения синонимии (предпочтения, условной эквивалентности). Из множества ЛЕ класса условной эквивалентности только одной ЛЕ придается статус дескриптора, другие становятся запрещенными к использованию при индексировании документов. Это синонимы, омонимы, термины, обладающие многозначностью (лексической полисемией, например сыворотка молока, сыворотка крови). Между ЛЕ устанавливаются ассоциативные связи – любой вид смысловых отношений между понятиями, возможный в данной предметной области, кроме отношений синонимии. Например:

Растения

а вечнозеленые растения

а декоративные растения

а с.-х. растения

а ядовитые растения

Ассоциативно связанные с дескриптором термины имеют пометку а.

4. Оформление лексики ИПЯ. На последнем этапе упорядочения производится оформление терминосистемы в виде нормативного словаря

наиболее правильных терминов, который является ее результатом. Тезаурус как терминологический справочник нормализованных и рекомендуемых терминов может считаться стандартом отраслевой терминологии и это является основной задачей нормализации и гармонизации отечественной терминологии. При этом под гармонизацией мы понимаем приведение в систему, в состояние гармонии, т.е. порядка и единообразия отечественной терминологии. Но чаще термин гармонизация употребляют в отношении межъязыковых терминологических отношений. В целях межъязыкового упорядочения терминологий – гармонизации – производится сопоставление отраслевых терминосистем (тезаурусов) двух языков – русского и английского. Основное средство гармонизации – корректировка содержания и форм национальных терминов с целью установления между ними точных соответствий. Результаты гармонизации оформляются в виде нормативных переводных словарей, в том числе стандартов. В нашем случае результатом гармонизации зарубежной, международной (ФАО) и национальной терминологии является ИПТ ЦНСХБ, в котором каждому основному термину, разрешенному к использованию и получившему статус дескриптора на русском языке, подбирается эквивалент на английском языке и эквивалент в международных тезаурусах, таких как AGROVOC (тезаурус AGRIS, ФАО ООН), CABI (Великобритания) [6]. Это дает возможность ученому сравнить выражения понятия на обоих языках, а также использовать тезаурус при написании научных статей и переводе их на английский язык. Вообще в процессе формирования словарной статьи ИПТ, выстраивания иерархических деревьев и определения статуса термина происходит изучение и сравнение данной терминологической области и отдельного термина в международных и зарубежных тезаурусах (AGROVOC и тезауруса базы данных CABI международного информационного бюро стран Британского содружества), что также способствует гармонизации терминологии.

Также в целях гармонизации отечественной и зарубежной отраслевой терминологии была создана русскоязычная версия международного многоязычного тезауруса AGROVOC международной базы данных AGRIS, в которой включено некоторое количество национальной терминологии. Русскоязычную версию можно использовать при составлении поискового запроса в международной базе данных AGRIS и других англоязычных базах данных по сельскому хозяйству и продовольствию.

Заключение

В процессе нормализации отраслевой научной лексики создаются иерархические деревья термина, выявляются его парадигматические связи, создается словарная статья на новые понятия. Результатом нормализации новой отраслевой лексики является обогащение контента ИПТ иерархическими деревьями, составляющими словарные статьи новых терминов.

Таким образом, тезаурус является средством нормализации и гармонизации отраслевой терминологии. Использование тезауруса в качестве терминологического справочника способствует внедрению стандартизированной, нормализованной, унифицированной лексики.

Список использованной литературы

1. Хакиева З. У. Место терминологии в лексической системе языка // Современная филология: материалы междунар. науч. конф. (г. Уфа, апрель 2011 г.). — Уфа: Лето, 2011. — С. 209-212.
2. Тезаурус: что это такое. Словарь тезаурус, который больше, чем словарь. <http://fb.ru/article/147602/tezaurus-chto-eto-takoe-slovar-tezaurus-kotoryiy-bolshe-chemslovar>
3. Гендина Н. И. Информационно-поисковые тезаурусы: основные виды и области применения // Научные и технические библиотеки. — М.: Государственная публичная научно-техническая библиотека России, 2008. — С. 5-14.
4. ГОСТ 7.25-2001. Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления (введен в действие с 1 июля 2002 г.). — М.: ИПК Издательство стандартов, 2002. — 14 с.
5. Харченко Л. Т., Пирумова Л. Н., Костин А. К., Соколова Ж. В. Система семантических отношений между терминами Информационно-поискового тезауруса по сельскому хозяйству и продовольствию // Аграр.наука, 2015. — № 4. — С. 36-38.
6. Голубева Е. А., Курченко Г. А., Милевская И. А., Пирумова Л. Н. Проблемы адекватности перевода терминов отечественного тезауруса на английский язык // Матер. XVI междунар. науч. конфер. "Биб. дело – 2011: Библиотечно-информ. деятельность в услов. модернизации об-ва". — М., 2011. — С. 442.-445.

L. N. Pirumova

Federative State Budgetary Scientific Institution
"Central Scientific Agricultural Library", Moscow

NORMALIZATION AND HARMONIZATION OF SECTORAL TERMINOLOGY IN THE BRANCH THESAURUS: GOALS AND OBJECTIVES