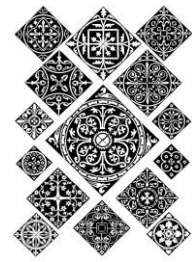


СЕКЦИЯ 7. НАЦИОНАЛЬНАЯ ЛОКАЛИЗАЦИЯ КОМПЬЮТЕРНЫХ СИСТЕМ И ТЕРМИНОЛОГИЯ



ПРЕИМУЩЕСТВА АВТОМАТИЗАЦИИ ТЕРМИНОГРАФИЧЕСКОЙ РАБОТЫ

*А. Гурбанова, Институт Информационных
Технологий НАНА Баку, Азербайджан, afruz1961@gmail.com*

***Аннотация:** в статье рассматриваются понятия «лексикография» и «терминография», их история, теоретические и практические особенности. Отмечена необходимость разработки единых принципов и методов составления словарей, описаны необходимые для этого условия. Определены перспективные направления лексикографии и терминографии, отмечены преимущества их компьютеризации. Проанализированы процессы автоматизации работ по лексикографии и терминографии и отмечена необходимость создания терминологической базы данных, позволяющей накопление и быструю обработку больших объемов терминологической информации. Показаны возможности терминологической базы данных, созданной в рамках Национальной терминологической информационной системы Азербайджана.*

***Ключевые слова:** лексикография, терминография, терминологический словарь, компьютерная лексикография, компьютерная терминография, терминологическая база данных.*

ADVANTAGES OF TERMINOGRAPHIC AUTOMATION

*A.Gurbanova, Institute of Information Technology of ANAS,
Baku, Azerbaijan, afruz1961@gmail.com*

***Annotation:** the article reviews the concepts of "lexicography" and "terminography", their history, theoretical and practical features. The need to develop common principles and methods for compiling dictionaries is noted, and the necessary conditions for this are described. Prospective directions of lexicography and terminography are determined, and the advantages of their computerization are noted. The processes of automation of works on lexicography and terminography are analyzed. The necessity of building terminological database enabling the accumulation and rapid processing of large volumes of terminological information is noted. The opportunities of terminological database established within the framework of the National Terminological Information System of Azerbaijan are shown.*

***Key words:** lexicography, terminography, terminological dictionary, computer lexicography, computer terminography, terminological database.*

Введение

На сегодняшний день информационные технологии являются неотъемлемой частью лингвистики, как и любой другой сферы профессиональной деятельности. Использование компьютеров и прикладных программ в лингвистических исследованиях, переводах и изучении языка сегодня имеет большое значение. Первым шагом к

использованию информационных технологий в лингвистике был переход из печатной и устной формы в цифровую.

При определении соответствующих разделов лингвистики, связанных с использованием информационных технологий, необходимо обратить внимание на различия между теоретической и прикладной лингвистикой.

Теоретическая лингвистика – это раздел лингвистики, изучающий становление государственного языка, его историю, соответствие законам. Прикладная лингвистика, развивающаяся с 20-х годов прошлого столетия, является разделом лингвистической науки, специализирующейся в области разработки методов решения практических проблем, связанных с использованием языка.

Будучи разделом прикладной лингвистики, вычислительная лексикография имеет целью создание компьютерных словарей, разработку программного обеспечения для лингвистической базы данных и поддержки лексикографических работ. Основными направлениями традиционной и вычислительной лексикографии являются определение структуры словарей и места лексических статей, а также разработка принципов проектирования различных типов словарей (Щипицина Л.Ю., 2013).

Одним из перспективных направлений вычислительной лексикографии и прикладной лингвистики является работа над электронными терминологическими словарями и терминологическими базами.

Терминография, входящая в состав лексикографии, занимается разработкой специальных терминологических словарей. Терминография тесно связана с терминоведением, наукой об особой лексической единице языка. Соответственно, вычислительная терминография – это наука о разработке электронных терминологических словарей.

В данной работе рассматриваются перспективы развития лексикографии, связанные с автоматизацией работ терминографии, являющейся одной из её ветвей.

1. Лексикография и терминография

Искусство разработки словарей имеет давнюю историю, но как научная дисциплина с её концептуальным аппаратом и методами исследования лексикография была сформирована в двадцатом веке. Так, начиная с 70-х годов прошлого столетия лексикография, выйдя из сферы разработки словарей, перешла в отдельный научно-практический предмет, появилась теория работы над словарями. В научной литературе лексикография определяется как один из разделов лингвистики, специализируется в теоретической и практической разработке словарей, выполняет функции общественной важности: изучение языка, представление и нормализация языка, межкультурное и внутриязыковое общение, изучение языка с научной точки зрения (Бобунова М. А., 2009).

В истории мировой лексикографии известны такие выдающиеся лексикографы как Владимир Даль (Россия), Е. Grimm, К. Grimm и Г. Пауль (Германия), Э. Литтре и П. Ларусс (Франция), И. Малорни (Италия), В. Вартбург (Швейцария), Х. Вебстер (США) (Комарова З. И., 2016).

Как было отмечено выше, современная лексикография традиционно делится на 2 части (Дамбуев И.А., 2011). К ним относятся:

- теоретическая лексикография, которая изучает методы разработки словарей и составные элементы словарей;
- практическая лексикография, которая исследует типологию словарей и осуществляет классификацию словарей по отдельным формам (исторические, фразеологические, этимологические и др.).

Начиная с 80-х годов прошлого столетия начал формироваться один из разделов лексикографии, а именно терминография. Следует отметить, что в работах учёных в научной литературе, связанной с проблемами лексикографии и терминографии, даётся неоднозначная интерпретация данных терминов.

Известный испанский лексикограф Хулио Касарес в своей книге «Введение в современную лексикографию» отмечает, что лексикография – это искусство разработки словарей. (Касарес Х., 1958).

Терминография – это интегративный научно-практический предмет по истории, теории и практике терминологических словарей. В него включены методология, методы, методика и технология для оптимального проектирования, разработки и использования терминологических словарей, созданных для решения различных научных и практических вопросов (Комарова З. И., 2016).

Терминологические словари имеют немаловажное значение для всестороннего налаживания связей на международном и внутригосударственном уровнях, а также для создания и обеспечения информационных служб и систем (Герд А. С., 2005).

По мнению российского лингвиста А. С. Герда, «терминография» и «лексикография» имеют одно и то же значение. Объект научно-технической (или терминологической) лексикографии состоит из практики и теории создания специальных (терминологических) словарей (Герд А. С., 1996).

Термин «терминография» выведен из словосочетания «терминологическая лексикография». Терминография занимает одно из приоритетных направлений терминоведения (Лейчик В. М., 2006).

Терминологическая лексикография или терминография – это наука о проектировании, разработке и использовании словарей специализированной лексики (Гринёв-Гриневиц С. В., 1993).

По мнению российского филолога К.Я.Авербуха, функция создания терминологических словарей традиционной лексикографии по-прежнему **использует** свои основные правила, но материал, используемый разработчиками терминологических словарей, поставленные задачи и методы их решения привели к появлению нового научного направления – терминографии (Авербух К. Я., 2004).

В статье азербайджанского языковеда А.М.Гурбанова представлена главная цель и назначение лексикографии, заключающиеся в сборе, включении в систему слов из словарного состава какого-либо языка, а также фразеологических единиц, в объяснении их происхождения, значения и принадлежности (Гурбанов А.М., 2003).

Многие из проблем, изучаемых терминологами, возникают на практике в ходе разработки терминологических словарей, и решение этих проблем влияет на методы составления словарей. В то же время изучение любой сферы специальной лексики связано с терминографией, поэтому, результаты работ по обнаружению, исследованию и регулированию терминологии обычно составляются в форме словаря. Терминография тесно связана с терминоведением, поскольку в рамках терминоведения решаются такие проблемы, как многозначность терминов, принятие синонимов и омонимов, выбор их эквивалентов в других языках. В связи с этим многие терминоведы считают, что терминоведение является теоретической базой терминографии или же терминография является одним из разделов терминоведения (Гринева С.В., 1995).

Как известно, словари играют важную роль в жизни людей. Они, являясь инструментом для изучения родного и иностранного языка, способствуют расширению запаса знаний и укреплению существующих знаний. Сегодня они играют большую роль и в ускоренном развитии науки и техники. Следует отметить, что в последнее время

число терминологических словарей стремительно растёт, что приводит и к увеличению числа других словарных форм.

В 1920-1952 гг. в Азербайджане количество терминологических словарей в 10 раз превышало количество других словарных форм. С 50-х годов прошлого века по сегодняшний день издано до 300 двуязычных, многоязычных и толковых словарей (Садыгова С., 2011).

Поскольку область знаний объективна, а термины и терминологические системы привязаны к конкретному языку, важным вопросом терминографии является стандартизация и унификация терминов, а также их однозначная интерпретация в различных языках.

Унификация терминологических систем основывается на терминологических стандартах. В настоящее время в мире существует более 20 000 стандартов, связанных с организацией терминологических систем. Помимо этого, существуют терминологические стандарты разных уровней – международного, государственного и даже уровня отдельной компании или фирмы. В связи с этим, вопрос об унификации терминологии и терминологических систем также должен быть неотъемлемой частью государственной и локальной языковой политики (Баранов А.Н., 2007).

В соответствии со всем вышеизложенным, перспективными направлениями современной лексикографии и терминографии являются:

- проектирование, внедрение и эффективная методология использования различных типов словарей для разных целей, реализация методики и технологии;
- выявление новых проблем теории лексикографии и терминографии;
- утверждение новых научных парадигм, интегрированных с лексикографией и терминографией;
- совершенствование существующих видов и жанров словарей;
- создание новых типов словарей, основываясь на новые принципы и проблемы;
- информатизация и компьютеризация работы над словарями для более ускоренного создания словарей, а также упрощения работы лексикографа / терминографа.

2. Автоматизация лексикографии и терминографии

Известны преимущества компьютеризации лексикографии и одного из её разделов – терминографии: компьютер может быстро предоставить информацию о синонимах и антонимах слова, цитатном материале и т.д.

Задачи, стоящие перед современной лингвистикой, ставят использование компьютерных технологий на первое место.

Здесь можно выделить нижеследующие направления:

- лингвистическое обеспечение различных типов информационных систем;
- машинный перевод;
- разработка систем, понимающих естественный язык;
- разработка системы использования информации, состоящей из звукового речевого сигнала и т.д.

Ещё в 70-х годах прошлого века были отмечены возможности применения компьютера в работе лексикографии (Марчук Ю.Н., 2007). Таким образом, компьютер может выстроить в алфавитном порядке лексические единицы, а также выполнять более тяжёлую работу, которая потребовала бы у лексикографов затраты большего времени и труда.

Современные вычислительные системы позволяют автоматизировать лексикографическую работу практически на каждом этапе – от выбора цитат до редактирования и печати словаря. Автоматизация такой работы, широкое использование

компьютерного программного обеспечения увеличивает производительность труда лексикографов (терминографов). В результате формируется новое направление – лексикографическая информатика. Лексикографическая информатика включает в себя создание автоматических словарей, а также разработку программ поддержки лексикографической работы.

Основные аспекты лексикографической информатики нижеследующие:

- автоматическое извлечение различных словарей из текстов с помощью компьютерных инструментов;
- теоретические и практические аспекты составления компьютерных словарей для системы обработки естественного языка;
- создание и реализация машинной версии традиционных словарей.

Одной из актуальных проблем является включение популярных словарей и запросов в машинные носители и создание на их основе новых словарей.

В статье (Rundell M., Hanks P., Schryver G.M., 2015) описаны три разных метода создания и сбора лингвистических данных, применяемых для создания словаря: контент пользователя; модель вики; модель краудсорсинга. При этом отмечено, что все три метода — важный потенциал для лексикографии и терминографии.

Одним из ключевых вопросов терминографии является создание сложных автоматизированных терминологических словарей.

Таким образом, терминологическая информатика будет развиваться по двум основным направлениям:

- создание печатных словарей на основе компьютерной технологии;
- создание электронных словарей, действующих только в памяти компьютера или на магнитных носителях.

Известно, что в настоящее время осуществляется переход от печатных словарей к электронным. Именно за счёт электронных словарей произошло ускоренное развитие теории и практики лексикографии и терминографии, и в первое десятилетие XXI века уже существовало 600 типов электронных словарей, доступных для 40 языков (Дубичинский В. В., 2009). Основными типами электронных словарей являются (Гринёв-Гриневиц С. В., 2008):

1. информационно-поисковый тезаурус, который служит информационно-поисковой системе, используя при этом языки поиска информации;
2. база данных;
3. база терминологических данных;
4. база знаний;
5. база терминологических знаний;
6. машинные корпусы и др.

Повышение эффективности поиска связано с повышением качества информационно-поисковых тезаурусов. По этой причине были разработаны методы оптимизации информационно-поисковых тезаурусов, которые позволяют выбрать более оптимальный лексикон для тезауруса [Мамедова М.Г., Скороходько Э.Ф., 1985).

Некоторые исследователи полагают, что «...компьютеризация работы над словарями определяет будущий характер лексикографии. Благодаря компьютерной обработке лексикографических данных разработка словарей будет продолжаться стремительным шагом, возможно, XXI век назовут Золотым веком лексикографии» (Дубичинский В. В., 2009).

В статье (Kosem I., Gantar P., Logar N., Krek S., 2014) были представлены два различных лексикографических и терминографических исследовательских проекта. В обоих проектах использовался один и тот же метод автоматического выведения данных

корпуса и были представлены аналогичные и отличительные особенности полученных данных. Даны перспективы автоматизации в лексикографии и терминографии.

3. Терминологическая база данных

Как было упомянуто выше, в цели уменьшения трудоемкости создания словарей в последнее время были разработаны автоматизированные методы. Автоматизация лексикографической работы привела к необходимости создания терминологической базой данных (ТБД). Современные компьютерные технологии позволяют обрабатывать и хранить в ТБД большое количество терминов из различных областей знаний.

Создание ТБД требует больших затрат, и замена его состава является с крупномасштабной работой. Поэтому очень важно предварительно изучить содержание терминологической информации ТБД. Необходимость организации структуры и автоматической обработки терминологической информации в больших количествах сделала эту область актуальной в терминографии.

ТБД включает и себе следующие функции:

- предоставление информационных и справочных услуг специалистам в различных областях знаний;
- обеспечение традиционного перевода научно-технической литературы;
- обеспечение системы машинного перевода;
- лингвистическое обеспечение автоматизированных информационных систем;
- обеспечение работы по терминологическому регулированию;
- подготовку и публикацию терминологических словарей;
- унификацию определённых терминов;
- подготовку научных отчётов по языку.

В настоящее время создаются ТБД по различным областям знаний и практической деятельности.

В Оттавском университете Канады разработали концептуальную основу Терминологической базы знаний (ТБЗ). Реализация концепции ТБЗ нацелена на облегчение процесса приобретения знаний (Meyer I., Skuce D., Bowker L., Eck K., 1992).

ТБД России есть во Всероссийском научно-исследовательском институте классификации, терминологии и информации по стандартизации и качеству (ВНИИКИ Госстандарта России) (www.vniiki.ru). Крупнейшая в мире ТБД, принадлежащая немецкой фирме "Siemens", содержит 2,5 миллиона терминологических записей на 8 языках. В Германии ТБД есть в Институте стандартизации и других руководящих органах. Большая международная ТБД была создана в Люксембурге в Бюро терминологий Европейской комиссии (www.iate.europa.eu). Созданная под названием «Eurodicautom» ТБД функционирует на нескольких языках. ТБД были созданы и используются во Франции, Канаде, Швеции, Бельгии и других странах мира. У Международной организации по стандартизации есть гигантская ТБД (www.iso.org/obp/ui).

Следует отметить, что в Азербайджане разработана концепция Национальной терминологической информационной системы (НТИС) (Rasim M. Alguliyev, Afruz M. Gurbanova, 2018). В рамках НТИС создан Национальный терминологический веб-портал с целью сбора терминологических словарей, разработанных в различных областях в единой информационной системе (www.terminologiya.az).

Таким образом, ТБД имеют большой потенциал развития, они стали инструментом для проведения серьезных научных исследований на основе автоматизированного хранилища, а также имеют большое практическое значение.

Вывод

В данной работе рассмотрен вопрос определения перспектив развития, касающихся автоматизации терминографических работ, и были получены следующие результаты:

- установлено, что основными задачами современной терминографии являются создание надёжной классификации специализированных словарей, анализ спроса на отдельные формы словарей, анализ способов повышения их качества, изучение методов отбора и организации терминологической информации в словарях и методов автоматизации терминологической деятельности.

- автоматизация терминографических работ даст возможность решить следующие проблемы:

- обеспечение отдельного подхода к разработке терминологических словарей;
- предоставление выбора оптимальной последовательности для определения ключевых характеристик при разработке словарей;
- предоставление возможности многочисленным специалистам, занимающимся разработкой словарей, для организации своей работы и объективного оценивания результатов.

- создание в Азербайджане ТБД способствует повышению эффективности терминотворчества и поможет решить нижеследующие задачи:

- моделирование терминологической системы азербайджанского языка;
- создание общетеоретических и общенаучных тезаурусов;
- исследование азербайджанской терминологии.

ЛИТЕРАТУРА:

1. Авербух К. Я. Общая теория термина, Москва: Издательство МГОУ, 2004, 252 с.
2. Alguliyev R.M., Gurbanova A.M. The Conceptual Foundations of National Terminological Information System, International Journal of Education and Management Engineering(IJEME), 2018, pp.31-49
3. Баранов А. Н. Введение в прикладную лингвистику, Учебник; МГУ, Филол. фак., Изд. 3-е., Москва: Изд-во ЛКИ, 2007, 358 с.
4. Бобунова М. А. Русская лексикография XXI века, Учеб. Пособие, М., Флинта : Наука, 2009, 200 с.
5. Всероссийский научно-исследовательский институт классификации, терминологии и информации по стандартизации и качеству, <http://www.vniiki.ru>
6. Герд А. С. Прикладная лингвистика. СПб., СПбГУ, 2005, 268 с.
7. Герд А. С. Научно-техническая лексикография, Прикладное языкознание. СПб., Изд-во СПб, 1996, 287 с.
8. Гринёв-Гриневиц С. В. Введение в терминоведение, М., Москва, Лицей, 1993, 230 с.
9. Гринева-Гриневиц С.В. Введение в терминографию, М., Изд-во МПУ, 1995, 158 с.
10. Гринёв-Гриневиц С. В. Терминография, Терминоведение, Учеб. Пособие, М., Академия, 2008, 304 с.
11. Гурбанов А.М. Современный азербайджанский литературный язык. I том. Баку, «Нурлан», 2003, 450 с.
12. Дамбуев И.А. Современная лексикография: статус и направления развития, Бурятия, Россия // Вестник Бурятского Государственного Университета, 2011, стр. 16-21.
13. Дубичинский В. В. Терминография и стандарты, Лексикография русского языка : учеб. Пособие, М., Флинта: Наука, 2009. 432 с.
14. IATE — European Terminology Database, <http://www.iate.europa.eu>
15. International Organization for Standardization, Online Browsing Platform, <http://www.iso.org/obp/ui>
16. Касарес Х. Введение в современную лексикографию, Изд-во иностранной литературы, 1958, 354 с.
17. Kosem I., Gantar P., Logar N., Krek S. Automation of Lexicographic Work Using General and Specialized Corpora: Two Case Studies, Proceedings of the XVI EURALEX International Congress: The User in Focus, Lexicography and Language Technologies, 2014, pp. 355-364.
18. Комарова З. И. Когда и почему возникла терминография как научно-прикладная дисциплина? // Актуальные проблемы германистики, романистики и русистики : сб. науч. Работ, Екатеринбург, Изд-во УрГПУ, 2016, ч. 3, с. 90—107.
19. Лейчик В. М. Терминоведение (Предмет, методы, структура), М., КомКнига, 2006, 256 с.

20. Мамедова М.Г., Скороходько Э.Ф. Метод оптимизации информационно-поискового тезауруса // НТИ, сер. 2 – Информационные процессы и системы, 1985, стр. 12–17 (Москва).
21. Марчук Ю.Н. Компьютерная лингвистика: учеб. Пособие, М., АСТ: Восток-Запад, 2007, 317 с.
22. Meyer I., Skuce D., Bowker L., Eck K. Towards a New Generation of Terminological Resources: An Experiment in Building a Terminological Knowledge Base, 1992, COLING '92 Proceedings of the 14th conference on Computational linguistics, Vol. 3, p. 956-960, <http://acl.ldc.upenn.edu/C/C92/C92-3146.pdf>
23. Rundell M., Hanks P., Schryver G.M. Crowdsourcing, wikis, and user-generated content, and their potential value for dictionaries // International Handbook of Modern Lexis and Lexicography, 2015, pp. 1-16.
24. Садыгова С. Терминология азербайджанского языка. Баку, «Элм», 2011, 380 с.
25. Терминологическая Комиссия при Кабинете Министров Азербайджанской Республики, <http://www.terminologiya.az>
26. Щипицина Л.Ю. Информационные технологии в лингвистике : учеб. пособие, М., ФЛИНТА: Наука, 2013, 128 с.
27. Afruz M. Gurbanova
28. Institute of Information Technology of ANAS, Baku, Azerbaijan
29. afruz1961@gmail.com



ТРЕХЪЯЗЫЧНЫЙ ТЕРМИНОЛОГИЧЕСКИЙ СЛОВАРЬ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ: ЦЕЛИ, МЕТОДОЛОГИЯ, ПЕРСПЕКТИВЫ

*Д. А. Киселев, О. Я. Юсупов, Самаркандский государственный
институт иностранных языков, Самарканд, Узбекистан,
dkiselyov@umail.uz, otabekuz10@mail.ru*

***Аннотация:** в статье излагаются цели и методические основы создания трехязычного (англо-узбекско-русского) терминологического словаря по компьютерной лингвистике, а также перспективы его использования. Работа над составлением словаря ведется группой преподавателей СамГИИЯ в рамках проекта ERASMUS+ CLASS, направленного на создание междисциплинарной магистерской программы по компьютерной лингвистике для вузов Узбекистана.*

***Ключевые слова:** компьютерная лингвистика, тюркские языки, терминологический словарь, дефиниция, толкование.*

TRILINGUAL DICTIONARY OF COMPUTATIONAL LINGUISTICS' TERMINOLOGY: AIMS, METHODOLOGY, AND PERSPECTIVES

*D. A. Kiselev, O.Ya.Yusupov, Samarkand state institute of foreign languages,
Samarkand, Uzbekistan, dkiselyov@umail.uz, otabekuz10@mail.ru*

***Annotation:** the article presents aims and methodological basis of compiling of trilingual (English-Uzbek-Russian) dictionary of computational linguistics' terminology as well as perspectives of its use. The work on the dictionary is being carried out by the workgroup of Samarkand State Institute of Foreign Languages in the framework of ERASMUS+ CLASS project which aims the design of an interdisciplinary MA program on Computational linguistics for Uzbek universities.*