

ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ

УДК 81'33

Жданович Марья Артуровна
преподаватель кафедры теории
и практики перевода №1
Минский государственный
лингвистический университет
г. Минск, Беларусь

Marya Zhdanovich
Lecturer at the Translation & Interpreting
Theory and Practice Department № 1
Minsk State Linguistic University
Minsk, Belarus
zma@tut.by

КОМПОНЕНТНЫЙ АНАЛИЗ ПО СЛОВАРНЫМ ДЕФИНИЦИЯМ
КАК СПОСОБ СИСТЕМАТИЗАЦИИ ТЕРМИНОЛОГИИ
(на материале предметной области «Информационные технологии»)

COMPONENT ANALYSIS OF DICTIONARY DEFINITIONS
AS A WAY OF TERMINOLOGY SYSTEMATIZATION PROCESS
(Based on the «Information Technology» Domain)

Статья посвящена проблеме систематизации терминологии в рамках решения задачи разработки компьютерного словаря предметной области «Информационные технологии» (ИТ). Приводятся результаты компонентного анализа отобранных 919 однословных терминов предметной области ИТ, направленного на выявление понятийных компонентов и установление парадигматических отношений между ними. Обосновывается определение 38 систематизирующих кластеров исследуемой предметной области для построения классификационной схемы понятий и создания лингвистической базы данных разрабатываемого компьютерного терминологического словаря.

Ключевые слова: термин; систематизация; понятийный компонент; кластер; моделирование; компьютерный словарь; схема понятий.

The article is focused on the problem of terminology systematization as a part of the development of a computer dictionary on Information Technology. The results of the component analysis of 919 extracted one-word lexical units of the IT domain are given. The analysis is aimed at identifying conceptual components and establishing paradigmatic relationships between them. 38 systematising clusters of the IT domain are defined to create a classification scheme of concepts and to develop a linguistic database for a computer dictionary of terminology.

Key words: term; systematization; conceptual component; cluster; modelling; computer dictionary; scheme of concepts.

Разработка компьютерных терминологических словарей в настоящее время приобретает особую актуальность и востребованность. Как отмечают специалисты в области лексикографии (С. В. Гринев, Ю. Н. Марчук, В. Д. Табанакова), чтобы получить качественный словарь, необходимо провести предварительную систематизацию терминологии предметной области. Анализ ряда исследований в области терминографии позволил установить, что для эффективной систематизации терминологии активно используется моделирование как метод лингвистических исследований.

В современной прикладной лингвистике актуальным направлением компьютерной терминографии является тезаурусное моделирование, отличительная особенность которого заключается в представлении лексического состава языка в виде систематизированных групп слов. Моделирование тезауруса как способа систематизации терминологии предполагает реализацию трех основных этапов [1, с. 126–127]: установление родовидовых, партитивных и ассоциативных отношений между терминами, построение иерархической классификационной схемы понятий, состоящей из кластеров, и распределение терминологической лексики между кластерами.

В отличие от общепринятых видов тезаурусов, которые отображают иерархические, синонимические, антонимические и другие отношения между лексическими единицами, в разрабатываемом компьютерном словаре ставится задача отобразить системность терминологии на основе общности понятийной структуры терминов. Такой способ представления лексики способствует формированию терминологической компетенции у современного специалиста и повышению качества профессиональной коммуникации посредством предоставления информации о понятийной структуре терминов, общих понятийных компонентах между терминами и их кластерной принадлежности. Кроме того, разрабатываемый компьютерный словарь позволит проводить автоматическую обработку специальных текстов с целью визуализации их понятийного содержания с помощью облака терминов, а также автоматическое определение места новых терминов в понятийном пространстве предметной области.

Представленное исследование по систематизации терминологии в рамках решения задачи разработки компьютерного словаря проводилось на материале терминологии предметной области «Информационные технологии» (ИТ), которая характеризуется стремительным темпом развития, что влечет за собой регулярные изменения ее терминологической составляющей. Материалом для исследования послужил 821 однословный термин, отобранный методом сплошной выборки из стандарта «Information technology – Vocabulary»: ISO/IEC 2382:2015 [2], и 98 терминов, извлеченных посредством метода семантико-статистического анализа из созданного корпуса текстов научных статей современных англоязычных журналов, изданных в период с 2016 по 2019 годы, а также из корпуса текстов руководств по эксплуатации новейшего программного обеспечения компании Microsoft за 2019 год.

Решение поставленной задачи по систематизации терминологии с помощью метода тезаурусного моделирования предполагает на первом этапе построение иерархической классификационной схемы понятий, состоящей из кластеров. На основе изученных работ установлены два основных подхода к разработке схемы понятий некоторой предметной области: ономаσιологический, т.е. с опорой на внеязыковую, или экстралингвистическую картину предметной области [1; 3], и семасиологический, т.е. с опорой на анализ

понятийного содержания отдельных терминов выбранной предметной области [4]. Применение второго способа представляется объективным в силу двух обстоятельств: во-первых, при ономаσιологическом подходе отмечается влияние мировоззрения автора словаря на общую схему тезауруса, а во-вторых, описание языковых феноменов с помощью заранее заданных логических схем существенно огрубляет знания об изучаемом феномене [5, с. 25].

В рамках семасиологического подхода для анализа понятийного содержания терминов при создании иерархической классификационной схемы как способа представления понятийного пространства предметной области ИТ в компьютерном терминологическом словаре наиболее целесообразным представляется использование метода компонентного анализа на основе словарных дефиниций. Следует отметить, что среди лингвистов нет единства взглядов относительно методики выделения содержательных компонентов значения. По мнению Т. З. Зевахиной, методика реализации компонентного анализа зависит от целей исследования и может варьироваться [6]. Ряд исследователей, например Ю. Д. Апресян, А. Вежбицкая, Ч. Дж. Филлмор, А. М. Кузнецов и др., опираются на идею о том, что язык может быть описан с помощью небольшого набора компонентов значения и допускают высокую степень обобщения на основе анализа собственного опыта и представлений об исследуемых объектах. Принимая во внимание, что данное исследование направлено на анализ терминологической лексики, а не общеупотребительной, необходимо учитывать, что одно из основных отличий термина от нетермина заключается в том, что термин – это лексическая единица, обозначающая специальное понятие [3, с. 7]. Следовательно, в составе терминологической системы предметной области термины через систему понятий репрезентируют объективную действительность. Поэтому в проводимом исследовании предлагается такая методика реализации компонентного анализа, которая позволила бы установить компоненты значений на основе анализа понятийного содержания терминов исключительно с опорой на специальные словари и построить классификационную иерархическую схему предметной области, которая может быть положена в основу организации лексики в компьютерном терминологическом словаре. При предлагаемом методе систематизация терминов осуществляется на основе интегрирующих понятийных компонентов.

В представленном исследовании компонентный анализ проводился для решения двух задач: для выявления понятийных компонентов терминов и для создания иерархической классификационной схемы предметной области на основе полученной понятийной структуры.

В целях решения первой задачи по установлению наиболее полного понятийного содержания терминов представляется целесообразным использовать метод обобщения словарных дефиниций, который позволяет установить максимальное количество понятийных компонентов [7]. В качестве

источника дефиниций отобраны шесть новейших англоязычных толковых терминологических словарей предметной области ИТ с 2002 по 2016 г. издания [8–13]. Кроме того, проанализированы дефиниции, содержащиеся в стандарте «Information Technology – Vocabulary»: ISO/IEC 2382:2015 [2]. Таким образом, из семи источников методом обобщения словарных дефиниций отобраны 6319 дефиниций исследуемых 919 лексических единиц (табл. 1).

Т а б л и ц а 1

Результаты отбора дефиниций из терминологических словарей¹

Источник дефиниций	Количество дефиниций
Information technology. Vocabulary : ISO/IEC 2382:2015(en) [2]	963
Oxford Dictionary of Computer Science [8]	704
Dictionary of Computing [9]	846
Dictionary of ICT [10]	1018
Dictionary of Computer and Internet Terms [11]	479
Microsoft Computer Dictionary [12]	869
Dictionary of Computer and Internet Terms [13]	1440
Всего	6319

На основе анализа работ, посвященных проблеме реализации метода обобщения словарных дефиниций и с учетом прикладного характера проводимого исследования целесообразным представлялось определить следующие этапы выделения понятийных компонентов отобранных 919 лексических единиц для анализа их понятийной структуры на основе дефиниций:

1) определить все специальные понятия, обозначаемые каждым отобранным термином предметной области ИТ;

2) выделить из дефиниций лексические единицы, которые непосредственно отображают понятийное содержание термина; в представленном исследовании данные единицы обозначены как вербальная фиксация понятийного содержания термина, которая включает в себя формулировку необходимых и достаточных признаков понятийного содержания терминов [14, с. 150];

3) осуществить поиск элементов вербальной фиксации понятийного содержания терминов в толковых словарях с целью установления их принадлежности к понятийному пространству исследуемой предметной

¹ Например, термин *interface* зафиксирован во всех словарях. В стандарте он имеет одну дефиницию, в словарях [8–10] – по 5 дефиниций, в словаре [11] – 1 дефиницию, в словаре [12] – 3 дефиниции, а в словаре [13] – 8 дефиниций.

области. В случае, если элемент вербальной фиксации обозначает специальное понятие исследуемой предметной области, выделить на его основе понятийный компонент;

4) на основе выделенных понятийных компонентов определить понятийную структуру отобранных терминов.

Проиллюстрируем этапы выделения понятийных компонентов на примере термина *memory*. Данный термин представлен в стандарте [2] и в шести словарях [8–13], в каждом из которых зафиксировано одно значение в следующих дефинициях:

1) *all of the addressable storage space in a processing unit and all other internal storage that is used to execute instructions* [2];

2) *a device or medium that can retain information for subsequent retrieval* [8];

3) *storage space in a computer system or medium that is capable of retaining data or instructions* [9; 10];

4) *the space within a computer where information is stored while being actively worked on* [11];

5) *a device where information can be stored and retrieved* [12];

6) *any temporary storage space used within or in conjunction with a computer* [13].

На основе компонентного анализа выявлена следующая понятийная структура термина *memory*: *device, processing unit / central processing unit, storage, computer system, computer, to store, to obtain*.

Проведенный таким образом компонентный анализ 6319 дефиниций позволил установить понятийную структуру отобранных 919 терминов и выделить 755 интегральных понятийных компонентов значений, которые послужили основой моделирования.

В целях решения второй задачи компонентного анализа по определению кластеров для построения классификационной схемы понятий были установлены парадигматические отношения между выявленными 755 понятийными компонентами. Как отмечается в [1, с. 54–55], в терминологии фундаментальными являются иерархические отношения, что подтверждается результатами проведенного анализа: 71 % из 755 понятийных компонентов определяется посредством родовидовых дефиниций и 16 % – посредством партитивных дефиниций. В незначительном количестве также установлены квазиродовые (2,3 %), описательные (4,1 %) и трансформационные (6,8 %) дефиниции.

Поскольку среди анализируемых дефиниций преобладают родовидовые и партитивные дефиниции, устанавливались холонимы и меронимы наивысшего порядка в рамках выбранной предметной области. В результате проведенного анализа родовидовых и партитивных отношений получено 278 понятий наивысшего порядка. Например, для понятий, входящих в структуру термина *memory*, установлены следующие гиперо-гипонимические ряды:

- *storage* → **device** → *any printer, storage, display, input, or output mechanism*
- *central processing unit* → *computational and control unit of a computer* → *computer* → **device** → *any printer, storage, display, input, or output mechanism*
- **computer system** → *central processor with storage and associated peripherals*
- *computer* → **device** → *any printer, storage, display, input, or output mechanism*
- **to store** → *to enter or retain information*
- *to retrieve* → **to obtain** → *to get or receive*

В данных рядах конечными терминами, которые не дефинируются посредством явно выраженного родового термина предметной области ИТ, являются *device, computer system, to store, to obtain*, следовательно, они представляют собой гиперонимы наивысшего порядка в рамках ИТ.

В целях объединения гиперонимов и холонимов в кластеры, необходимо определить неиерархические отношения между ними. В результате проведенного исследования по установлению неиерархических отношений между выявленными понятиями наивысшего порядка определено количество понятийных компонентов, объединенных такими семантическими связями как действие – инструмент (26), действие – объект (16), действие – результат (14), действие – состояние (8), место – действие (10), объект – пространство (4), пространство – действие (2), свойство – объект (4), способ – действие (10), условие – действие (2), а также отношениями синонимии (10) и антонимии (2).

На основании установления парадигматических отношений между 755 понятийными компонентами выявлено 38 систематизирующих кластеров, которые вместе с входящими в их состав понятийными компонентами образуют иерархическую классификационную схему предметной области ИТ (табл. 2).

Т а б л и ц а 2

Кластеры, образующие понятийное пространство предметной области ИТ

Кластер	Кол-во понятийных компонентов	Кластер	Кол-во понятийных компонентов
hardware	109	condition / state / status	11
data / information	97	person	10
to alter / to change	56	statement	10
structure	39	to check	10
quantity	35	to store	9
language	34	reasoning	8

software	25	to operate	8
to change / to move	24	to protect	8
telecommunications	22	manipulation / to manipulate	5
to connect / to link / connection / link	17	problem	5
action / event	15	access	4
characteristic	15	computer science	3
types of data / information	15	time	3
computer graphics / graphics	14	process / task	2
file	14	to search	2
system	14	to assign / to define	2
to arrange	14	to choose	2
to create	12	to enter / to type	2
to represent	12	to use	1

Таким образом, второй этап тезаурусного моделирования, или автоматическое распределение терминов между кластерами, реализуется посредством понятийных компонентов, которые составляют их понятийную структуру. В качестве примера приведем проанализированный ранее термин *memory*.

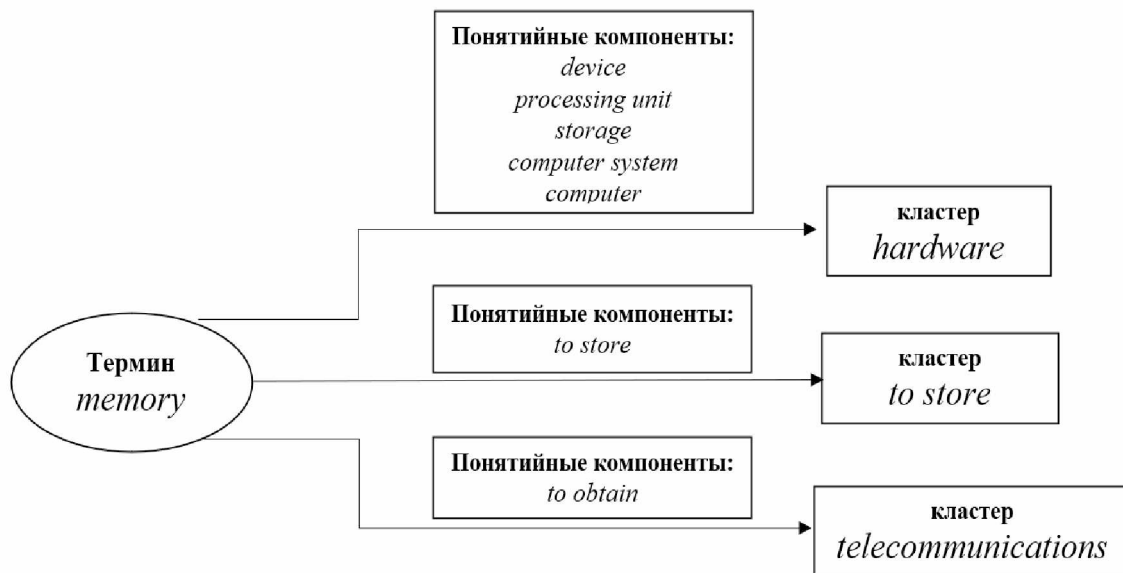


Рис. 1. Место термина *memory* в иерархической классификационной схеме понятий

Следует отметить, что третий этап тезаурусного моделирования, или автоматическое установление отношений между терминами, также реализуется посредством понятийных компонентов. В представленном исследо-

вании связанными на понятийном уровне считаются термины, имеющие в своей понятийной структуре общий понятийный компонент. Например, посредством понятийного компонента *to store* термин *memory* связан с терминами *archive, disk, memory, to compress* и т.д.

Заключение. В рамках создания компьютерного терминологического словаря предметной области ИТ задача по систематизации терминологии для ее репрезентации в словаре решена методом моделирования. Для создания модели предметной области на основе метода компонентного анализа 6319 словарных дефиниций установлены:

- 1) понятийная структура 919 терминов, содержащая 755 понятийных компонентов;
- 2) семантические отношения между 755 компонентами;
- 3) 38 систематизирующих кластеров, которые составляют классификационную схему понятий.

Понятийные компоненты, семантические отношения и кластеры позволили нам разработать макро- и микроструктуру компьютерного терминологического словаря, основной функцией которого является репрезентация системности терминологии на уровне текста и на уровне предметной области.

ЛИТЕРАТУРА

1. *Комарова, З. И.* Моделирование двуязычного словаря-тезауруса по экономике : монография / З. И. Комарова, А. А. Прошина. – Екатеринбург : Форум-книга, 2009. – 276 с.
2. *Information technology. Vocabulary* : ISO/IEC 2382:2015(en) [Electronic resource] // Online Browsing Platform (OBP). – Mode of access : <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en>. – Date of access : 06.01.2021.
3. *Табанакова, В. Д.* Идеографическое описание научной терминологии в специальных словарях : автореф. дис. ... д-ра филол. наук : 10.02.21 / В. Д. Табанакова ; Тюменский гос. ун-т. – СПб., 2001. – 38 с.
4. *Караулов, Ю. Н.* Общая и русская идеография / Ю. Н. Караулов. – Изд. стер. – М. : ЛИБРОКОМ, 2014. – 355 с.
5. *Баранов, А. Н.* Семантика фразеологизмов: иерархия или сеть? / А. Н. Баранов, Д. О. Добровольский // Компьютерная лингвистика и интеллектуальные технологии : по материалам ежегод. междунар. конф. «Диалог–2010», Бекасово, 26–30 мая 2010 года; редкол.: А. Е. Кибрик (гл. ред.). – Вып. 9 (16). – М. : Изд-во РГГУ, 2010. – С. 25–29.
6. *Зевахина, Т. С.* К типологии методик компонентного анализа / Т. С. Зевахина // Прикладные аспекты лингвистики / МГУ им. М. В. Ломоносова, Филол. фак. : М. : Изд-во МГУ, 1989. – С. 26–47.
7. *Стернин, И. А.* Методы описания семантики слова / И. А. Стернин. – Изд. 2-е. – М., Берлин : Директ-Медиа, 2015. – 34 с.

8. *Butterfield, A.* Oxford Dictionary of Computer Science / A. Butterfield, G. E. Ngondi, A. Kerr. – Oxford : Oxford University Press, 2016 – 627 p.
9. *Collin, S. M. H.* Dictionary of Computing / S. M. H. Collin. – 5th ed. – London : Bloomsbury, 2004. – 345 p.
10. *Collin, S. M. H.* Dictionary of ICT / S. M. H. Collin. – 4th ed. – London : Bloomsbury, 2004. – 288 p.
11. *Downing, D. A.* Dictionary of Computer and Internet Terms / D. A. Downing [et al.]. – 10th ed. – New York : Barron's, 2009. – 560 p.
12. *Microsoft Computer Dictionary* – 5th ed. – Redmond : Microsoft Press, 2002. – 637 p.
13. *Rigdon, J. C.* Dictionary of Computer and Internet Terms: in 2 vol. / ed. J. C. Rigdon. – Cartersville : Eastern Digital Resources, 2016. – 1 vol.
14. *Шелов, С. Д.* Очерк теории терминологии : состав, понятийная организация, практические приложения / С. Д. Шелов. – М. : ПринтПро, 2018. – 472 с.

Поступила в редакцию 08.02.2023