

УДК 004.82: 519.766.24: 519.766.48: 81.13

Гусева А.Е.

Московский государственный областной университет

СТРУКТУРА БАЗЫ ЗНАНИЙ В КОГНИТИВНОЙ ЛИНГВИСТИКЕ*

A. Guseva

Moscow State Regional University

KNOWLEDGE BASE STRUCTURE IN COGNITIVE LINGUISTICS

Аннотация. В статье определяется понятие базы знаний. Устанавливаются типы и виды знаний применительно к когнитивной лингвистике. Приводятся сведения о структуре базы знаний, разработанной автором для исследования фразеологической картины мира немецкого и русского языков. База знаний позволила провести качественный анализ фразеологической базы данных, включающий изучение парадигматических отношений фразеологических единиц, их стилистических и национально-культурных особенностей, разработку лексикографической кодификации исследуемого материала на основе универсальной модели.

Ключевые слова: база знаний, база данных, когнитивная лингвистика, концепт, фразеологическая картина мира, лингвокогнитивное моделирование, лексико-фразеологическое поле, универсальная идентификационная модель, корреляционный анализ.

Abstract. The article contains the definition of the term "knowledge base". It is concerned with knowledge types and forms with respect to cognitive linguistics. The article provides information on the knowledge base structure developed by the author to analyze German and Russian pictures of the world. The database enabled to perform a quality analysis of the phraseology databases including the study of paradigmatic relations of phraseological units, their stylistic and national and cultural particulars as well as the development of lexicographical codification of the material under study based on the universal model.

Key words: knowledge base, database, cognitive linguistics, concept, phraseological picture of the world, linguistic and cognitive modeling, lexical and phraseological field, universal identification model, correlation analysis.

Понятие «база знаний» впервые возникло в экспертных системах [3] как результат эволюции понятия «база данных» и связывалось с тем фактом, что формализация информации в виде базы данных оказалась недостаточной для принятия решений в экспертных системах. Оказалось, что полная и упорядоченная информация в любой области знаний требует соответствующих этой области знаний методов обработки, т. е. оперирования данными по определенным правилам с целью выработки всевозможных вариантов решения поставленной задачи. В любой экспертной системе принятие решения – это выбор наилучшего варианта решения задачи из множества возможных вариантов, выработанных при помощи базы знаний. Таким образом, базу знаний можно определить как совокупность баз лингвистических данных, универсальной идентификационной модели, формализующей лингвистическую информацию, а также правил и методов оперирования данными с помощью формализованных процедур с целью выработки максимально возможного множества вариантов решения поставленной задачи.

В прикладной лингвистике базы знаний были разработаны, в первую очередь, для систем машинного перевода (СМП), которые работали как особый класс экспертных систем [10]. В последние годы разработчикам СМП удалось разработать процедуры обработки лингвистической информации, которые позволили добавить к европейским и некоторые восточные языки, например китайский, а также реализовать работу онлайн-переводчиков за счет гибридной технологии перевода, совмещающей классический, основанный на правилах, метод машинного перевода, и статистические методы [16].

Целью настоящей статьи является рассмотрение структуры базы знаний применительно к задачам когнитивной лингвистики.

* © Гусева А.Е.

Как известно, в рамках когнитивной лингвистики язык рассматривается как общий когнитивный механизм, исследующий те когнитивные структуры, которые свойственны человеку как *homo loquens* [1, 9; 12].

Понимание языка как «функциональной знаковой системы, предназначенной для извлечения, хранения и передачи информации» [13], вызвало необходимость изучения представления знания в языке. Наименование фрагментов действительности языковыми формами предполагает комплексное исследование процессов и механизмов когнитивной деятельности человека посредством самого языка, в частности, системного описания и объяснения механизмов усвоения языка, изучения принципов структурирования этих механизмов [8, 21].

Основное содержание когнитивной лингвистики определяют многочисленные специальные понятия, к которым относятся термины лингвистического характера: знание как система знаний на основе имеющейся информации или как система знаний, полученных в результате познавательной деятельности человека; категоризация; концептуализация и ее основная единица – концепт; языковая и концептуальная картина мира; прототип; стереотип и т. п.

Исходя из основной задачи данной статьи, остановимся подробнее на вопросе о типах и видах знаний в лингвистике, а затем рассмотрим процесс формирования базы знаний для задач когнитивной лингвистики и проиллюстрируем его на примере полученных нами научных результатов.

В рамках когнитивной науки знания квалифицируются как результаты постижения закономерностей объективного мира или как адекватное отражение реальности в сознании человека [4, 58].

Современная когнитивная наука (КН) выделяет различные типы и виды знаний, которые можно представить в виде следующих оппозиций [14, 284-285]: 1) *эмпирическое* (основанное на полученном опыте) – *рациональное* (логическое) знание; 2) *языковое* (вербальное) – *экстралингвистическое* (невербальное) знание; 3) *декларативное* (внешнее, словесное (например, строение человеческого тела) – *процедурное* (например, знание, как водить машину, как пользоваться метро и т. п.).

В состав базы знаний входят следующие компоненты: 1) *языковые знания*: а) знание языка – грамматики (с фонетикой и фонологией), дополненное знанием композициональной и лексической семантики; б) знания об употреблении языка; в) знание принципов речевого общения; 2) *внеязыковые знания*: а) о контексте и ситуации, знания об адресате; б) общефоновые знания (то есть знания о мире); знания о событиях, состояниях, действиях и процессах и т. д. [2, 7].

Для современной когнитивной лингвистики (КЛ), в частности для построения и понимания текста любого вида и типа особую важность приобретают следующие пять фундаментальных типов систем знания:

- 1) языковое знание (знание грамматики и лексики языка);
- 2) энциклопедическое знание;
- 3) знание мира (языковое + энциклопедическое знание = интеракциональные знания);
- 4) знания, полученные в результате коммуникации и межличностных отношений (*Interaktionswissen*) [21, 208];
- 5) лингвокультурологические знания.

Каждая из пяти вышеперечисленных систем знания, в свою очередь, может быть подвергнута дальнейшей структуризации. Развернём данное положение.

1. Языковое знание или знание того или иного языка предполагает владение, прежде всего, его лексикой и грамматикой. Оно предполагает преобразование ментальных представлений знания мира в языковые репрезентации.

Языковое знание охватывает правила грамматики и лексикон, являющийся своего рода «переключателем» между языковым знанием и знанием мира. Следует различать задачу реконструкции того, что соответствует «внутри нас» языку, от того, с одной стороны, в каком виде мы реконструируем языковую способность и знание языка, и, с другой стороны, какой статус мы приписываем этому знанию. При этом грамматика не отделена от лексики, а, наоборот, «синкретична с ней», лексикон же организован по сетевому принципу, он «представляет собой ассоциативно-семантическую сеть с включённой в неё и в значительной мере лексикализованной грамматикой» [11; 13].

Лексикон содержит в себе интериндивидуальное знание, которым обладают члены

языкового социума в результате познания социальной и естественной окружающей действительности. Семантические репрезентации данных лексикона (слов, фразеологизмов) выступают также в роли своеобразных «переключателей», осуществляющих доступ к энциклопедическому знанию (ср. известное аналогичное противопоставление А.А. Потебни «ближайшее – дальнейшее значение слова» [15]).

2. Энциклопедическое знание.

3. Знание мира.

Разделение знания мира на языковое знание и энциклопедическое знание не обусловлено психологической реальностью, а имеет исследовательско-методологические причины, является по большей части эвристическим, то есть основанным на осознании чего-либо методом познания окружающей действительности.

4. Интеракционные знания представляют собой комплексную систему знаний, учитывающую следующие аспекты:

- знания о целях, которые преследуются теми или иными языковыми высказываниями; знания об условиях, при которых эти цели могут быть достигнуты; знания о последствиях, которые могут возникнуть после высказывания;

- знания об общих/глобальных организационных схемах (знания о типах текстов); знания о коммуникативных нормах.

Из перечисленных аспектов особую роль играет знание коммуникативных целей или иллюкативное знание (т. е. интерпретация речевого акта с точки зрения его коммуникативной функции, например: вопрос, побуждение и т. п.) [20, 208-209].

5. Лингвокультурологические знания. В процессе усвоения человеком того или иного языка он становится носителем данной целостной смысловой системы, членом данного этнокультурного сообщества. Посредством культуры человек включается не только в социальное поле своей жизнедеятельности, но и в систему общечеловеческих форм существования, обусловливаемых константами всеобщей культуры, в том числе и знаниями. При этом знание становится базисной формой когнитивной организации культуры, поскольку оно представляет собой важный фактор упорядочения

повседневной жизни и деятельности человека. Именно благодаря своей актуальности и важности в его жизнедеятельности знание становится элементом культуры.

Одновременно с этим знание является продуктом культуры, поскольку процесс освоения человеком правил, норм и стереотипов мышления, выработанных предшествующими эпохами, происходит путем включения знаний в реальную историческую практику людей, принадлежащих к соответствующему этнокультурному сообществу.

Таким образом, лингвокультурологическая ценность категории «знание» определяется тем, что оно конституируется в формах человеческой субъективности, организованной обычаями, традицией и культурой [5, 73; 19].

Особую систему знаний образуют человеческая память, ментальный (индивидуальный) лексикон и концептуальные структуры. При этом в сознании человека следует выделить следующие когнитивные области, требующие пристального внимания специалистов разных наук:

- языковая компетенция как способность производить и понимать определенные содержания;

- конкретное знание как результат познавательной деятельности;

- когнитивная деятельность (в узком смысле слова – как осознание и определение смысла определенного высказывания, действия и т. п.) [14, 287].

Далее возникает вопрос о системах представления знаний к когнитивной лингвистике, в частности – вопрос о соотношении языковых категорий и когнитивных структур. Уже в 1987 г. Дж. Лакоффом была выдвинута гипотеза о том, что языковые категории определяются и формируются когнитивными структурами, которые он называет «идеализованные когнитивные модели» (ИКМ – *idealized cognitive models*) [17].

По мнению Дж. Лакоффа, когнитивный аппарат, которым обладает человечество, охватывает 4 макромоделли или 4 когнитивные структуры:

1) пропозициональные модели (модели предложений), с помощью которых из ситуаций вычленяются категориальные элементы,

их свойства, и между ними устанавливаются определенные отношения;

2) схематичные модели-картинки, отражающие определенные представления, которые формируют категориальные классы, например: «самолет» как некая труба с крыльями, «человек» как некая фигура с головой, туловищем, руками и ногами и т. п.;

3) метафорические модели, позволяющие идентифицировать какую-либо абстрактную область и сравнить ее с конкретной, эмпирически доступной областью/денотатом;

4) метонимические модели, основанные на логическом соотношении, например: «часть-целое», позволяющие перенести характеристику одного элемента на все множество предметов.

В современной КЛ, в частности в когнитивной семантике, выявлены и сформулированы различные формы представления знаний в виде некоторых ментальных структур. В центре научного исследования процесса познания, бесспорно, находится «концепт» как ментальная информационная единица в долговременной памяти человека, элементарный краеугольный камень познания [18, 378-379] и его основные разновидности:

- *фрейм* – типизированная ситуация и ее языковая реализация, например: «покупки» – *покупать, продавать, продавец, цена, товар, оплачивать* и т. п.;

- *сценарий (скрипт)* – стереотипная последовательность эпизодов, событий в их динамике (ср. «стадион» как фрейм: *спортсмен, зритель, соревнование*; понятия же «посещение стадиона», «игра в футбол», «прогулка», «экзамен» – будут уже являться скриптами, или сценариями);

- *гештальт* (термин Х. ф. Эренфельда) – единая функциональная структура, объединяющая эмоциональные рациональные элементы, например: *любовь, учёба в университете* и т. п.;

- промежуточное положение между представлением и понятием занимает *схема* (например, схематично нарисованная ребёнком фигурка) [8].

Одним из примеров формирования и реализации базы знаний, в частности при исследовании «фразеологической картины мира»,

может служить проведённое нами лингвокогнитивное моделирование лексико-фразеологических полей (ЛФП) немецкого и русского языков на базе основных коммуникативных концептов ментальной деятельности человека: мыслительная/слуховая/зрительная/речевая деятельность [6; 7].

Под *концептом* будем понимать сложное структурно-смысловое ментально-перцептивное, как правило, лексически и/или фразеологически вербализованное образование, которое обладает следующими релевантными признаками:

1) целостность единой понятийной основы (ср.: зрительная / слуховая / мыслительная / речевая деятельность),

2) наличие культурной ценности,

3) возможность функционального замещения человеку в процессе коммуникации предметов мира (в широком смысле слова), являющихся результатом его речевой, зрительной, слуховой и мыслительной деятельности,

4) наличие понятийных, образных и ценностных компонентов,

5) полевая структура.

Каждый из вышеозначенных концептов определяем как *макроконцепт* (ядро), каждые из последующих его модификаций являются *микрoконцептами/микрoполями*, также включающими в себя понятийный, образный и ценностный компоненты [7, 60-61].

База знаний при лингвокогнитивном моделировании исследуемых ЛФП имеет следующую структуру.

I. База данных экстра- и интралингвистического характера. Так, экстралингвистическая характеристика исследуемого материала включает концептуальную характеристику основных процессов человеческой деятельности, определение которых имеет существенное значение для определения состава и логических признаков идентификации исследуемых ЛФП через призму четырёх обозначенных концептов.

База данных лингвистического характера сформирована в результате лингвистического моделирования каждого ЛФП и складывается из следующих этапов.

1. Идентификация лексических (ЛЕ) и фразеологических единиц (ФЕ) на предмет при-

надлежности к ЛФП через призму соответствующего концепта: «слуховая/ зрительная/ мыслительная/ речевая деятельность», т. е. определение конституентного состава поля. В результате изучения словарных дефиниций некоторых наиболее авторитетных словарей немецкого и русского языков лексико-фразеологический корпус исследуемых ЛФП в немецком и русском языках составил соответственно:

- «Слуховая деятельность» - около 42 ЛЕ и 316 ФЕ, реализованных в более чем 3000 примеров;

- «Зрительная деятельность» - около 104 ЛЕ и 640 ФЕ, реализованных в более чем 4000 примеров;

- «Мыслительная деятельность» - около 96 ЛЕ и 1056 ФЕ, реализованных в более чем 6000 примеров;

- «Речевая деятельность» - около 81 ЛЕ и 2030 ФЕ, реализованных в более чем 8000 примеров. Таким образом, общее количество ЛЕ и ФЕ составило около 4365, реализованных в более чем 21000 примеров современных немецкого и русского языков.

2. Выявление эксплицитных и имплицитных идентификаторов принадлежности ФЕ к соответствующему ЛФП.

3. Установление структурно-семантической организации поля, т. е. выявление макрополей (МАП) на основе макроконцептов (МАК).

4. Выявление микрополей (МИП) на основе соответствующих микроконцептов (МИК).

5. Выявление центра и периферии ФЕ. Выявление параллельных связей между выявленными МАПами и/или МИПами.

6. Выявление межконцептуальных связей.

7. Структурно-синтаксическая и структурно-семантическая характеристика ФЕ в ЛФП.

II. Разработка универсальной структурной модели макро-и микрополей на базе основных коммуникативных концептов человека (слуховая / зрительная / мыслительная/речевая деятельность), ставшей непосредственным результатом формирования базы данных экстра- и интралингвистического характера. Модель включает 5 макрополей (МАП) и 29 микрополей (МИП) [7, 196].

III. Количественный анализ базы данных. Использование компьютерных технологий позволило оценить не только количественные ха-

рактеристики исследуемых явлений, но и подтвердить работоспособность, эффективность и достоверность разработанной универсальной модели структуры ЛФП, в частности, при более тонкой статистической обработке методами корреляционного анализа. Так, первичная статистическая обработка отражает частоту попаданий ФЕ в соответствующее подмножество для каждого коммуникативного концепта как в немецком, так и в русском языках. Табличный процессор программы Microsoft Excel позволил распределить исходное множество фразеологизмов по макро- и микрополям с использованием числовых маркеров «1,00», которые устанавливались в строке фразеологизма и столбце соответствующего макро- и микрополя, закодированного 29 маркерами, обозначенными $1/1, 1/2, \dots 5/3, 5/4$ [6, 23].

Для дальнейшей первичной статистической обработки были использованы возможности табличного процессора программы Microsoft Excel по определению состава подмножеств фразеологизмов, объединённых в соответствующие числовым маркерам микрополя. Для этих целей был использован режим фильтрации данных, вызываемый кнопкой *Автофильтр*, по условию значения численного маркера «1,00». Отфильтрованные табличным процессором фразеологизмы, принадлежащие, например, к микрополю с кодом $1/1$ занесены на лист 1-1 и так далее.

Таким образом, были сформированы необходимые для дальнейшей обработки восемь файлов табличного процессора программы Microsoft Excel: для ЛФП «Слуховая деятельность» – файлы SluchR.xls, SluchD.xls, для ЛФП «Зрительная деятельность» – файлы ZritR.xls, ZritD.xls, для ЛФП «Мыслительная деятельность» – файлы MyslR.xls, MyslD.xls и для ЛФП «Речевая деятельность» – файлы RechR.xls, RechD.xls, содержащие ФЕ русского и немецкого языков соответственно.

Полученные результаты использованы для более тонкой статистической обработки методами корреляционного анализа.

Использование компьютерных технологий, в частности, подсчёт коэффициента корреляции, который по всем анализируемым явлениям, как правило, стремится к величине +1, свидетельствует о сходности сущностных

распределений ЛЕ и ФЕ в сравниваемых языках при выражении важнейших человеческих способностей, связанных с коммуникацией, мышлением и мировосприятием окружающей действительности. Корреляционный метод применим, как известно, в случаях обнаружения и сопоставления, по крайней мере, двух рядов случайных распределений, характеризующих проявления исследуемого явления. Его суть состоит в выявлении тесноты взаимосвязи анализируемых рядов интересующих распределений. Количественные оценки значений коэффициентов корреляции свидетельствуют о чрезвычайно высокой идентичности языковых явлений в русском и немецком языках при условии применения методологии лингвокогнитивного моделирования ЛФП на базе систем МАП и МИП, выявленных с применением разработанной универсальной идентификационной модели.

Проведённые этапы статистической обработки обосновывают достоверность и надёжность последующих выводов непосредственно лингвистического анализа свойств коммуникативных концептов в свете сравнительного сопоставления изучаемых явлений в немецком и русском языках.

IV. Качественный анализ базы данных. Этот этап включает апробирование разработанной универсальной идентификационной модели макро- и микроструктуры исследуемых ЛФП на отдельных видах лингвистического анализа: при исследовании парадигматических отношений ЛЕ и ФЕ, их нормативно-стилистической характеристике, выявлении так называемых «культурных следов» в значении ФЕ – конститuentов ЛФП, поскольку фразеологический состав языка является наиболее прозрачным для воплощения средствами языка культуры данного народа.

V. Лексикографическая кодификация базы данных ЛЕ и ФЕ в виде идеографического словаря ЛЕ и ФЕ, обеспечивающего поиск ФЕ для выражения мысли ономасиологическим путём – «от понятия – к ФЕ».

В качестве основной идеи представления данных о ЛЕ и ФЕ использовалась идея о табулированном представлении данных, принятая в вычислительных и информационных системах, основным достоинством которого является

возможность постоянного развития и наращивания баз данных.

Словарная статья построена следующим образом: заглавное слово – название МАП соответствующего ЛФП и его номер.

Следующая информация отражена в табулированном виде с учётом шести параметров.

Шапка словарной статьи при табулированном представлении лексических и фразеологических единиц-компонентов МИП выглядит следующим образом:

МИП	ЛЕ	ФЕ	ЭИ	ИИ	ИП
1	2	3	4	5	6

Словарная статья предполагает, таким образом, наличие следующих сведений:

- 1) название и номер соответствующего микрополя;
- 2) ЛЕ – лексические единицы (существительные, прилагательные, глаголы), входящие в состав данного микрополя;
- 3) конкретная реализация ФЕ;
- 4) ЭИ – эксплицитный идентификатор, выраженный ЛЕ (существительным, прилагательным, глаголом), позволяющий отнести ФЕ к определённому МАП;
- 5) ИИ – имплицитный идентификатор, выраженный ЛЕ (существительным, прилагательным, глаголом, наречием), позволяющий отнести ФЕ к определённому МАП;
- 6) ИП – иллюстративные примеры.

Таким образом, становится очевидным, что когнитивная лингвистика решает глобальные задачи изучения разнообразных явлений языковой картины мира, поэтому каждая группа изучаемых явлений требует создания своей базы знаний. В частности, если поставить задачу изучения фразеологической картины мира с помощью базы знаний, необходимо создать базу данных, включающую массив фразеологизмов, относящихся к рассматриваемому явлению, затем построить идентификационную модель фразеологизмов в едином пространстве признаков (в нашем случае – это 29 маркеров), а затем на основе универсальной модели создать процедуры количественной и качественной обработки информации, выбираемой из базы данных фразеологизмов. Конечным итогом реализации базы знаний может стать

формализованный способ лексикографической кодификации фразеологизмов, вербально отражающих тот или иной концепт.

ЛИТЕРАТУРА:

1. Бабушкин А.П. Типы концептов в лексико-фразеологической семантике языка. Воронеж: Изд-во ВГУ, 1996. 104 с.
2. Герасимов В.И. На пути к когнитивной модели языка // Новое в зарубежной лингвистике: Когнитивные аспекты языка, вып. XXIII. М.: Прогресс, 1988. С. 5-11.
3. Голубева Л.Н. Когнитивные структуры экспертных знаний: методологический аспект // Экспертные системы. М.: Информприбор, 1990. Вып. 1. С. 15.
4. Горский Д.П. и др. Краткий словарь по логике. М.: Просвещение, 1991. 209 с.
5. Гумбольдт В. фон. Язык и философия культуры. М.: Прогресс, 1985. 450 с.
6. Гусева А.Е. Основы лингвокогнитивного моделирования лексико-фразеологических полей в немецком и русском языках: Автореф. дисс. ... д-ра филол. наук. М., 2008. 42 с.
7. Гусева А.Е. Основы лингвокогнитивного моделирования лексико-фразеологических полей в немецком и русском языках (Монография). М.: МГОУ, 2007. 363 с.
8. Демьянков В.З. Теория прототипов в семантике и прагматике языка // Структуры представления знаний в языке. М.: Институт научной информации по общественным наукам РАН, 1994. С. 32-86.
9. Жаботинская С.А. Концептуальная модель чатеречных систем: фрейм и скрипт // Когнитивные аспекты языковой категоризации. Рязань: Изд. РГПУ, 2000. С. 15-20.
10. Звегинцева О.В., Еникеев Б.Н. К построению ЭС для представления идеографических знаний // Экспертные системы. М.: Информприбор, 1990. Вып. 1. С. 22.
11. Караулов Ю.Н. Общая и русская идеография. М.: Наука, 1976. 364 с.
12. Кубрякова Е.С. Проблемы представления знаний в языке // Структуры представления знаний в языке. М.: ИНИОН РАН, 1994. С. 22-31.
13. Кубрякова Е.С. Язык и знание. М.: Языки славянской культуры, 2004. 555 с.
14. Ольшанский И.Г., Гусева А. Е. Лексикология. Современный немецкий язык = Lexikologie. Die deutsche Gegenwartssprache. М.: Academia, 2005. 415 с.
15. Потебня А.А. Мысль и язык. М.: Лабиринт, 1999. 268 с.
16. У компании PROMT резко выросли продажи в корпоративном сегменте // Бестселлеры IT-рынка. Аналитика российского рынка IT, 2011. № 1/ [Электр. источник]: <http://www.itbestsellers.ru/news/detail.php?ID=17706>.
17. Lakoff G. Women, Fire and Dangerous Things: What Categories Reveal about the Mind. Chicago: University of Chicago Press, 1987. – 614 p.
18. Metzler-Lexikon Sprache/Hrsg. von H. Glück. Stuttgart, Weimar: Metzler, 2000. 817 s.
19. Piirainen E. Phrasemes from a cultural semiotic perspective // Phraseologie: ein internationales Handbuch zeitgenössischer Forschung =Phraseology: an international handbook of contemporary research. – Hrsg. von/edited by H. Burger, D. Dobrovolskij, P. Kühn, N.R. Norrick. 1. Halbband / Volume 1. Berlin, New York: Walter de Gruyter, 2007. S. 208-219.
20. Techtmeier B., Motsch W., Viehweger D. Illocutionary knowledge and dialogue structure // S. Stati, E. Wiegand, F. Hundsnurscher (Hrsg.) / Dialoganalyse 111 Referate der 3. Arbeitstagung Bologna. Teil 1. Tübingen: Niemeyer, 1991. S. 197-209.
21. Viehweger D. Die Vielfalt textlinguistischer Forschungsansätze - methodologisches Dilemma oder notwendiger Pluralismus? // Linguistische Studien/Reihe A. Berlin, 1991. H. 209. S. 208.