

С.А. Осокина

Естественно-научные истоки лингвистической концепции тезауруса

S. A. Osokina

Natural Science Sources of the Linguistic Conception of Thesaurus

В статье раскрываются методологические основания некоторых положений естествознания, оказавших значительное влияние на современные лингвистические концепции соотношения языка и знания. Внимание фокусируется на формирующейся в настоящее время концепции тезауруса, предлагающей новый взгляд на сущность и способы изучения языковой системы знания. Цель статьи состоит в осмыслении концептуальной трансформации содержания понятия «тезаурус» в лингвистике под воздействием актуальных естественно-научных концепций методом критического анализа научной литературы по данному вопросу. Анализируются релевантные положения общей теории систем, теории информации, информатики и синергетики. В частности, рассматриваются свойства и функционирование сложных развивающихся систем, анализируется суть понятия «информация» в естествознании и изучается роль тезауруса как семиотической системы, необходимой для рецепции информации из окружающей среды и генерации новой информации. Производится сопоставление естественно-научных взглядов на семиотическую систему знания и соответствующих лингвистических идей, а также рассматриваются предлагаемые в естествознании методы исследования данной системы. Итогом работы является формулирование исходных положений лингвистической концепции тезауруса, сформированных в результате изучения рассмотренных естественно-научных теорий.

Ключевые слова: тезаурус, семиотическая система, информация, соотношение языка и знания, информатика, синергетика, языкознание.

DOI 10.14258/izvasu(2014)2.1-31

Данная статья продолжает рассматривать научные истоки формирующейся в лингвистике концепции тезауруса, которые мы начали освещать в работе 2011 г. [1].

Понятие «тезаурус» имеет довольно долгую историю своего развития в лингвистических трудах, поскольку соотносится с древнейшими представлениями

The article shows methodological foundations of some natural science statements which demonstrate significant influence on modern linguistic conceptions of the relationship between the language and the knowledge. The topic under discussion is a developing linguistic theory of thesaurus offering a new view on the semiotic system of knowledge. The purpose of the work is to understand the transformation of the linguistic notion «thesaurus» under the impact of the actual natural science conceptions by critical review of the papers at issue. The article analyzes the relevant ideas of the general theory of systems, the theory of information, informatics, or the computer science, and synergetics. In particular, we discuss properties and functioning of complex developing systems, analyze the natural science vision of information, and study the thesaurus as the semiotic system necessary for reception of information from the environment and generation of new information. We compare natural science views on the semiotic system of knowledge and corresponding linguistic ideas as well as consider the proposed by natural science methods of its study. The main result of the work is the formulation of the departure statements of the linguistic theory of the thesaurus formed as the consequence of the discussed natural science conceptions.

Key words: thesaurus, semiotic system, information, relationship between knowledge and language, computer science, synergetics, linguistics.

о системе слов языка и ее связи с системой понятий. Первые попытки представить слова языка в виде тезаурусных словарей, отражающих соотношение слов и понятийных категорий, относятся ко II–III вв. нашей эры, что означает становление понятия «тезаурус» за многие века до появления языкознания как науки. Современная лингвистика продолжает культивировать

традиционное представление о тезаурусе как о словаре особого типа, раскрывающем связь лексической системы языка с системой мыслительных категорий. Однако слово «тезаурус» в последнее время активно используется представителями других наук, и предлагаемые ими концепции не всегда соответствуют лингвистическим представлениям. Поскольку, по мнению ряда ученых, данное понятие обладает мощным эксплаторным и эвристическим потенциалом, связанным с возможностью объяснения особенностей строения мыслительного мира человека, сущности коммуникативных процессов и принципов жизнедеятельности информации в целом, назрела необходимость пересмотреть наполнение данного понятия в языкознании и сформировать лингвистическую концепцию тезауруса, отвечающую установкам современной методологии науки.

Цель данной статьи состоит в осмыслении концептуальной трансформации содержания понятия «тезаурус» в лингвистике под влиянием актуальных естественно-научных концепций методом критического анализа научной литературы по данному вопросу.

Обращение к понятию «тезаурус» в естественнонаучной мотивировано стремлением преодолеть некоторые проблемы изучения сложных информационных систем, начало исследованию которых было положено еще представителями общей теории систем в первой половине XX в. Господство системного подхода обусловило дальнейший ход развития методологии науки в целом.

Главной причиной возникновения системного подхода, провозглашающего изоморфизм в строении и функционировании различных объектов в природе (систем), является крушение механистического мировоззрения, доминировавшего с начала эпохи Нового времени и базирующегося на Ньютоновской «объективной» антропологии. Крушение связано прежде всего с формированием теории относительности и обращением пристального взгляда к фигуре субъекта исследования — ученому, чьи «объективные» научные выводы оказались весьма зависимыми от исходной позиции исследования. Именно тогда впервые стало очевидно, что, изучая природный объект, ученый взаимодействует с ним и не просто изменяет его, а является причиной ответных изменений в объекте: «Ведь сама система, обладающая собственным поведением, деятельностью, развитием и по своим «творческим» возможностям нередко не уступающая исследователю, — это не просто тот противостоящий исследователю и терпеливо ждущий своего отражения в его голове объект, который традиционно рассматривался в гносеологии» [2, с. 11].

Представители теории систем предлагают несколько вариантов определения системы. В качестве инварианта можно представить определение, включающее следующие параметры: система представляет собой

единство взаимосвязанных элементов, находящихся в определенных отношениях; статическое рассмотрение отношений между элементами складывается в структуру системы; система существует как таковая (идентична сама себе) до тех пор, пока ее структура сохраняется неизменной [3, с. 167].

Даже незначительные изменения структуры означают факт развития системы, которое возможно только при одном условии — взаимодействии со средой, при этом среда выступает как система более высокого порядка. В ходе взаимодействия со средой происходит обмен информацией между системой и средой.

Необходимо отметить важность понятия «информация», приобретшего особую актуальность в науке XX в. Хотя традиционно понятие информации (от лат. *informatio* — изложение, разъяснение) подразумевало осмысленное сообщение, главным образом словесное, знаковое, в теории систем информация понимается как физическая величина, «сравнимая по значению с понятием энергии в физике» [4, с. 31].

Таким образом, благодаря теории систем была создана почва для переосмысления понятия информации — появилась возможность поставить акцент на материальной стороне и деактуализировать смысловую сторону информации: информация — это в большей степени само сообщение, нежели его смысл. В течение XX в. эти идеи получили большой резонанс, особенно в семиотике и философии постмодернизма.

Выводы, сделанные представителями системного подхода, фундаментальны. Они создали не просто новую концепцию естествознания, но и новую методологию науки: само знание — это тоже система, и его можно изучать при помощи положений теории систем.

Идеи и методологические разработки, предложенные представителями общей теории систем, нашли свое дальнейшее развитие в ряде смежных дисциплин, среди которых теория информации, кибернетика, информатика, синергетика, а также теория игр и топология, включающая теорию сетей и теорию графов. Первые четыре оказали значительное влияние на развитие взглядов в лингвистике (например, в таких направлениях, как когнитивная лингвистика и прагмалингвистика) и поэтому требуют своего отдельного рассмотрения.

Основным понятием теории информации и кибернетики является «информация», понимаемая количественно как мера сложности структур и «канал связи» [5].

К. Шеннон первым стал рассматривать передаваемые сообщения с точки зрения математической статистики, изучая количество информации путем подсчета частоты появления букв алфавита в сообщении. Он не занимался исследованием содержания передаваемой информации, так как его интересовал сам механизм ее передачи. Предложенное Шенноном по-

нимание информации прочно закрепилось в точных науках и смежных отраслях.

В исследованиях по кибернетике, науке об управлении информационными системами посредством обратной связи, намечается выход на изучение содержания информации. Основатель кибернетики Н. Винер определяет информацию как «обозначение содержания, полученного из внешнего мира в процессе нашего приспособления к нему и приспособления к нему наших чувств» [6, с. 14]. Кибернетика не занимается изучением содержания информации, но включение в рассмотрение идеи обозначения — знаковости — способствует дальнейшему продвижению в этом направлении.

Изучением содержания информации более плодотворно занимается информатика. Это направление естественно-научных исследований представляет особый интерес в контексте настоящей работы, так как именно в лоне информатики возникла идея тезауруса как хранилища семантической информации.

В отличие от теории информации, в качестве информационной единицы информатика рассматривает не сигнал, а символ, т. е. двусторонний знак. Для адекватной оценки содержательной стороны информации представители информатики обращаются к семиотическим учениям о структуре знака. В частности, большое влияние оказала концепция Р. Карнапа, определяющего значение знака как его место в системе знаков, а величину семантической информации — посредством логической вероятности: чем больше логическая вероятность высказывания, тем меньше его содержание [7]. Идея измерения содержания информации путем измерения ее формы стала в информатике одной из основополагающих.

Формой существования и хранения семантической информации является тезаурус. Данная концепция наиболее последовательно развивается в трудах Ю. А. Шрейдера, определяющего тезаурус как формальную модель описания представлений о внешнем мире некоторого наблюдателя [8].

«Сложная система — пишут Ю. А. Шрейдер и А. А. Шаров, — имеет семиотическую (т. е. полноценно языковую) природу информационных связей между подсистемами в противовес системам, где имеется функциональная сигнализация» [9, с. 16]. Из семиотических учений воспринимается также мысль о трех аспектах существования системы знаков: семантическом, синтагматическом и прагматическом. Семантическая информация, передаваемая сложной системой, отождествляется с информацией прагматической, поскольку, независимо от того, сколько информации передано, воспринимается только та информация, которая необходима, ценна для принимающей системы.

Таким образом, семантическая модель передачи информации — тезаурус — необходима для того, что-

бы решить проблемы понимания информации. В шенноновской теории информации проблема восприятия не ставится, главное внимание уделено процессу передачи информации, а ее восприятие зависит от того, как настроен приемник, и предполагается, что он всегда правильно настроен. Однако это не так.

Тезаурус представляет собой семиотическую систему, на основе которой происходит восприятие информации, ее обработка и присвоение. «Речь идет о том, что одно и то же сообщение может по-разному восприниматься различными приемниками информации. Для описания подобной ситуации оказывается эвристически полезным представление о тезаурусе приемника, от состояния которого зависит возможность воспринять тот или иной фрагмент информации из данного сообщения и дать ему определенную интерпретацию в рамках данного тезауруса» [9, с. 118].

Тезаурус выступает как система репрезентаторов — способов представления сведений о действительности. Можно говорить о тезаурусе отдельного человека и целого коллектива. Тезаурус полностью определяет то, какие из поступающих сведений будут восприняты (и как они будут восприняты), а какие — нет (так как не укладываются в существующий тезаурус, не находят в нем соответствующих для себя форм).

Тезаурус задает систему семантических связей понятий. Каждое понятие в тезаурусе объясняется через набор других, т. е. характеризуется своим положением в тезаурусе. Фактически тезаурус человека — это «вербализованная, соотнесенная со словесным материалом совокупность его представлений о мире, включающая его познавательные установки» [9, с. 119]. Последние определяют ценность получаемой информации.

Итак, воспринимается только та информация, которая соотносится с имеющимся тезаурусом и немного превышает заданное в нем количество информации. То количество информации, которое превышает заданную в тезаурусе, является ценной информацией. Таким образом, обмен информацией между семиотическими системами происходит как получение ценной информации.

Из этого следует, что информационный обмен представляет собой процесс, телеологичный по своей сути. Если получаемая информация полностью совпадает с имеющимся тезаурусом, то она не несет никакой ценности. В таком случае, по мнению Шрейдера, можно говорить, что переданное количество информации равно нулю. Информация, определяемая как количество информации, — это только ценная информация. В этом смысле семантическая информация тождественна прагматической.

Как видно, при таком подходе имеется качественно иное основание (в отличие от идеи Шеннона) для понимания количества информации: его можно изме-

рять как изменение тезауруса воспринимающей системы: «При восприятии текста тезаурус человека может меняться. Изменение тезауруса можно записать алгебраически: равенство $b = x(a)$ означает, что субъект с тезаурусом a , восприняв текст x , изменяет свой тезаурус, превращая его в b » [9, с. 119].

Проблема состоит в том, что пока не известно, как можно измерить содержательную сторону тезауруса, гипотеза Шрейдера строится на очень важном допущении: «допустим, мы научились некоторым способом вычислять степень содержательности тезауруса» [9, с. 119]. Таким образом, Шрейдер не предлагает математической формулы для исчисления количества семантической информации, которая бы соответствовала формуле Шеннона для количества материальной информации. Однако это не мешает выстроить довольно последовательную в целом концепцию.

Трактовка тезауруса как системно организованной семантической информации в настоящее время широко развивается в синергетике, науке о самоорганизации системных процессов. Синергетика интегрирует точные, естественные и общественные науки путем обоснования изоморфизма эволюционного развития различных систем в природе и обществе при помощи математических методов исследования. Появление данного направления — также одно из закономерных следствий развития системного подхода, при котором на новом научном витке переосмысляются понятия системы и структуры, закономерности и случайности, развития и стабильности в более общих философских категориях порядка и хаоса.

Синергетика изучает открытые системы, способные к саморазвитию в процессе вещественного, энергетического и информационного обмена со средой.

Информация в синергетике понимается как случайный и запомненный выбор одного варианта из нескольких возможных и равноправных. Если представители общей теории систем и кибернетики пытались выстроить такую модель развития системы, при которой вероятность случайных событий не может привести к полному хаосу, так как предполагается наличие общей структуры ряда процессов, то представители синергетики, очевидно, не удовлетворившись идеей структуры, пытаются объяснить отсутствие хаоса наличием памяти системы. Система не делает абсолютно случайный выбор, просто не все причины или мотивы ее поведения объективно можно наблюдать. Объяснение, как это происходит, осуществляется при помощи изучения тезауруса системы.

Понятие «тезаурус» в синергетике также связано с семантической информацией, которая, как и в концепции Шрейдера, трактуется через категории цели и ценности: «Ценность информации зависит от цели, которую преследует рецептор» [10, с. 17].

В связи с представлением о ценной информации появляется представление об осмысленной инфор-

мации. В отличие от ценной информации, которая субъективна и связана с целями системы, осмысленная информация объективна и зависит от тезауруса. В традиционной информатике не ставится вопрос об эволюции ценной информации, поскольку цели развития информации определяются извне. В синергетике открытая самоорганизующаяся система способна порождать цели внутри себя (подобные идеи были намечены еще в общей теории систем). Поэтому рецепция информации, при которой система осуществляет выбор на основе имеющегося тезауруса, — это, с одной стороны, выбор самой системы, но с другой стороны — это выбор, навязанный извне, поскольку тезаурус — это то, чему система научается (приобретает) из внешней информационной среды. Система (в роли которой может выступать и человек, личность) делает выбор сама, но выбор, который она обречена сделать.

Тезаурус нужен не только для понимания информации (рецепции), но и для ее генерации. То, какая информация выбирается при рецепции, обуславливается тезаурусом; однако если выбор случаен — имеет место генерация новой информации. Иными словами, если в систему семантических связей, которая уже сложилась на данный момент, вносится новый элемент, это и есть генерация новой информации, и это и есть проявление случайности, поскольку происходит некоторая дестабилизация сложившихся связей, что не может быть целью самоорганизующейся системы, а может быть только случайным эффектом.

Тезаурус представляет собой иерархию ценностных уровней. Д. С. Чернавский иллюстрирует работу данной системы следующим примером: учась говорить, ребенок получает информацию от своих родителей — так у него складывается языковой тезаурус; повзрослев, человек овладевает определенной специальностью — появляется профессиональный тезаурус; дальнейшие ценностные выборы в жизни будут возможны только на основе прежних, более ранних, ведь выбор осуществляет не просто человек, владеющий языком, но и владеющий специальностью [10, с. 20–21].

Как ребенок научается языку? Как он приобретает свой самый первый тезаурус, на основе которого будет восприниматься уже вся последующая информация? — С точки зрения синергетики, переход к каждому последующему информационному уровню возможен только на основе уже имеющегося к данному моменту тезауруса, иначе не из чего будет сделать выбор. Следовательно, у ребенка к тому моменту, как он начинает учиться языку у родителей, уже должен иметься какой-то элементарный тезаурус? Представители синергетики говорят, что «необходимый для этого тезаурус у ребенка присутствует от рождения» [10, с. 21].

Данный вывод, несомненно, вызывает ассоциацию с учением о врожденности языковой способности Н. Хомского. Однако и в учении Хомского, и в цитируемой работе Чернавского идея врожденности языковых структур появляется там, где невозможно доказать, опираясь на достижения современной науки, откуда берутся эти исходные знания. По выражению Ю. Н. Караулова, идея о врожденности языковых структур «граничит уже с мистикой» [11, с. 7].

Следует отметить, что в большинстве естественно-научных исследований, имеющих отношение к языку, демонстрируются довольно общие представления о том, что такое язык: они, как правило, представляют собой знания о языке, свойственные рядовому его носителю, не владеющему системой знаний профессионального лингвиста.

Главным ограничением рассмотренных естественно-научных концепций является то, что они работают с математическими моделями, а не с реально существующими «вещественными» языковыми системами. Поэтому формулируемые ими положения не всегда соответствуют действительно происходящим языковым процессам.

Заслуживает критики и то, что стремление выявить универсальные принципы работы систем выливается в некоторых работах в редукцию сложных явлений и процессов.

Используемый, преимущественно математический, категориальный аппарат также не всегда поддается переложению на метаязык лингвистики и других гуманитарных наук, поскольку последний не имеет точных одно-однозначных значений.

Подводя итог рассмотрению естественно-научных концепций, имеющих отношение к изучению тезауруса, выделим те моменты, которые могут дополнить и изменить существующее в лингвистике представление о тезаурусе.

Тезаурус — это система знания.

В соответствии с общими свойствами систем тезаурус состоит из элементов и отношений между ними. Элементами признаются семиотические сущности, отношениями между ними — отношения на материальном и семантическом уровне. Связи между элементами можно изучать как структуру данной системы.

Основным принципом организации структуры тезауруса признается иерархический принцип, однако подчеркивается и важность горизонтального развертывания связей для определения специфики каждого уровня иерархии.

Развертывание тезауруса по горизонтали и вертикали создает сеть, обеспечивающую непрерывность информации, что подразумевает предположение о непрерывности перехода материальных связей в семантические (по крайней мере, не указывается, где проходит четкая граница между ними).

Тезаурус связывается с идеей поиска информации и памятью. Алгоритм поиска и память интерпретируются двояко — как материальные информационные процессы и семантические.

Для изучения информации предлагаются статистические методы исследования при анализе материальной стороны и логические — при анализе семантической стороны.

Тезаурус — это открытая система, следовательно, ему свойственны общие черты всех открытых систем. К ним относятся «поведение» как ответная реакция на воздействие внешней среды, осуществляемое, очевидно, по некоей структуре, хранящейся в памяти системы, вследствие чего это поведение можно в определенной степени контролировать и управлять им.

Тезаурус определяет цели и ценность семантической информации при обмене индивида знаниями со средой. Эта система необходима индивиду для рецепции входящей информации и генерации нового знания.

В качестве среды выступают тезаурусы других индивидов. Индивидуальный тезаурус связывает знания индивида и объективированные в виде семиотических сущностей знания, поэтому тезаурус имеется и в голове индивида, и как внешняя по отношению к индивиду информация. Тезаурусом также называется модель данной системы.

Информация — это физическая величина, соотносимая с понятием энергии, некая материальная субстанция. Информация передается от одного физического объекта к другому. Передача информации представляет собой передачу неких материальных вещественных форм. Поэтому информацию нужно изучать количественными методами.

Исследовать систему знания можно путем анализа изменения тезауруса реципиента информации.

Опираясь на данные идеи, можно сформировать новую лингвистическую концепцию тезауруса, значительно отличающуюся от других лингвистических концепций, исследующих языковые системы знания.

Наиболее последовательно языковая система знания изучается во множестве ответвлений когнитивной лингвистики, получая различные наименования, самые известные из которых — «картина мира» и «ментальный лексикон».

Принципиальным отличием разрабатываемой концепции является то, что тезаурус представляет собой неразделенную материально-семантическую сущность (в противовес «картине мира» и другим когнитивным системам, имеющим только семантическую природу), а также предлагается новый метод изучения этой системы.

Очевидным достижением рассмотренных естественно-научных концепций тезауруса является то, что они поднимают ряд методологически важных вопросов и предлагают нетрадиционные пути их ре-

шения, поэтому понятие «тезаурус» признается эвристичным и заслуживает глубочайшего изучения. В частности, заостряется проблема роли индивидуальных знаний субъекта при взаимодействии со сложными системами, что позволяет сгладить резкое противопоставление материального и идеального, а также поставить вопрос о зависимости идеального от материального (в то время как в лингвистике материальная сторона знака рассматривается как второстепенная сущность, лишь выражающая семантическое содержание).

Методологические трудности, с которыми сталкиваются естественно-научные концепции тезауруса, связаны, на наш взгляд, с тем, что не учитываются многие достижения лингвистики: несмотря на то что признается языковая сущность тезауруса, не используются лингвистические методы анализа и не обращается пристального внимания на многообразие проявлений языка.

Мы считаем, что данные трудности можно преодолеть при помощи правильного выбора единицы анализа — такой языковой единицы, которая удовлетворяла бы сетевому устройству тезауруса. Это не может быть отдельное слово или парадигматические классы слов, отражающие иерархии (что обычно применяется при построении тезаурусных словарей в лингвистике), это должны быть повторяющиеся синтагматические последовательности слов — устойчивые сочетания слов — реализующие иерархии в линейном пространстве.

Повторяемость данных последовательностей (регулярная воспроизводимость в речи) делает их полноправными единицами языка, а синтагматическая развертываемость заполняет горизонтальное пространство сети, приводя систему в движение, что позволяет изучать ее естественную жизнедеятельность.

С другой стороны, рассмотренные естественно-научные концепции позволяют преодолеть многие проблемы современных лингвистических исследований. В частности, с нашей точки зрения, эвристическим является использование количественных методов исследования, основывающихся на подсчете употребления определенных языковых выражений (т. е. частоты использования определенных материальных словесных последовательностей) при изучении семантической информации. Это позволит выявить реально работающие языковые механизмы, которые составляют актуальный тезаурус каждого носителя языка в отдельности и языковой тезаурус как сокровищницу живого знания в целом.

Итак, суть концептуальной трансформации содержания лингвистического понятия «тезаурус» под влиянием современных естественно-научных концепций сводится к формированию нового взгляда на сущность, структурное устройство и функционирование тезауруса, а также к разработке нового метода тезаурусного анализа языковых явлений с целью выявления принципов жизнедеятельности языковой системы знания.

Библиографический список

1. Осокина С. А. Лексикографические истоки лингвистической концепции тезауруса // Известия Алтайского государственного университета. — 2011. — № 2/2 (70).
2. Садовский В. Н. Задачи, методы и приложения Общей теории систем // Исследования по общей теории систем. — М., 1969.
3. Лоуссон И. А. Язык, информация и биологическая коммуникация // Исследования по общей теории систем. — М., 1969.
4. Бергаланфи Л. Общая теория систем — Критический обзор // Исследования по общей теории систем. — М., 1969.
5. Шеннон К. Работы по теории информации и кибернетике. — М., 1963.
6. Винер Н. Человек управляющий. — СПб., 2001.
7. Карнап Р. Философские основания физики. Введение в философию науки. — М., 1971.
8. Шрейдер Ю. А. Об одной модели семантической теории информации // Проблемы кибернетики. — М., 1965.
9. Шрейдер Ю. А., Шаров А. А. Системы и модели. — М., 1982.
10. Чернавский Д. С. Синергетика и информация (динамическая теория информации). — М., 2004.
11. Караулов Ю. Н. Русский язык и языковая личность. — М., 2007.