

**ТАШКЕНТСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ  
УНИВЕРСИТЕТ ПРЕПОДАВАТЕЛЬ КАФЕДРЫ  
УЗБЕКСКОГО ЯЗЫКА И ЛИТЕРАТУРЫ****Аширматова Манзура**

Возникновение теории Тезаурса

<https://doi.org/10.5281/zenodo.7924742>**АННОТАЦИЯ**

В данной статье рассказывается об общем понятие тезауруса, его изучение как системного знания в лексикографическом аспекте, а также в каких направлениях следует искать предпосылки возникновения теории тезауруса.

**АННОТАЦИЯ**

Ushbu maqolada tezaurusning umumiy xulosasi, uni tizimli bilim sifatida leksikografik jihatdan o'rganish, shuningdek, tezaurus nazariyasining paydo bo'lishi uchun zarur shart-sharoitlarni qaysi yo'nalishlarda izlash kerakligi tasvirlangan.

**АННОТАЦИЯ**

This article describes the general conclusion of the thesaurus, its study as a systemic knowledge in the lexicographic aspect, and also in what directions one should look for the prerequisites for the emergence of the thesaurus theory.

**Ключевые слова** тезаурус, сокровище, теория, идея, разбиение, система, инструмент, индивидуум, оболочка слова.

Хотелось бы нам рассказать молодым коллегам о возникновении и о понятии тезауруса.

Изучение роли языка в формировании знания является одной из ведущих задач современной лингвистики, особенно в русле когнитивной парадигмы. Однако, по выражению Р.М. Фрумкиной, «у нас пока нет теории, которую можно было бы, например, назвать «общей теорией знаковой поддержки получения и закрепления знания» [Фрумкина, 2001, с. 146].

Возникновение идеи построения теоретической концепции тезауруса в лингвистике не случайно. Такая идея могла появиться только на рубеже 2021 веков, и ее рождение подготовлено всем ходом развития научной мысли. Становление теории тезауруса в нач. 21 века – закономерно, своевременно и необходимо.

Для того чтобы определить, в каких направлениях следует искать предпосылки возникновения теории тезауруса, необходимо обратиться к наиболее общему определению ее ключевого понятия. Можно выделить следующие понимания тезауруса, имеющиеся в современной науке – лингвистическое, естественно-научное и более общее философско-методологическое, позволяющее использовать понятие «тезаурус» в междисциплинарных исследованиях.

В лингвистическом понимании тезаурус (от греч. thesaurus – сокровище, сокровищница) – это «словарь, в котором максимально полно представлены все слова языка с исчерпывающим перечнем примеров их употребления в текстах» или «идеографический словарь, в котором показаны семантические отношения (родовидовые, синонимические и др.) между лексическими единицами» [Шайкевич, 1998, с. 506].

Создание словаря, в котором были бы представлены все слова языка во всех контекстах своего употребления – задача, практически невыполнимая, поэтому такие тезаурусы существуют только для мертвых языков, например, «Thesaurus linguae latinae» или «Словарь польского языка XVI века». Тезаурус как идеографический словарь в идеале предполагает представление иерархической системы понятий, обеспечивающей поиск от смыслов к лексическим единицам, т.е. поиск слов, исходя из понятий [Караулов, 1981].

Однако составление подобного словаря также проблематично, уже с философско-методологических позиций, связанных с трудностью определению понятия и восстановления соотношения между словом и понятием, поэтому большинство существующих на сегодняшний день тезаурусов представляют собой словари, в которых отражены связи, имеющиеся у каждого отдельно взятого слова, в соответствии с современным представлением о системности лексики. В таких словарях семантические отношения слов передаются группировкой слов по тематическим рубрикам.

Усилия ученых по созданию тезаурусных словарей породили еще одно понимание тезауруса как «множество смысл выражающих единиц некоторого языка с заданной в нем системой семантических отношений». Данное определение сформулировано Ю. А. Шрейдером, автором работ по информатике, кибернетике и математической лингвистике [Шрейдер, 1976].

Тезаурус в таком понимании фактически определяет семантическую систему того или иного естественного языка, соотносимую с формализованным языком для автоматизированных систем управления. Такое представление мотивировало создание информационно-поисковых тезаурусов и проникновение понятия «тезаурус» в сферу естественных и точных наук.

Естественно-научное представление о тезаурусе базируется на определении последнего как информационно-поисковой системы, в которой каждой лексеме языка соответствует определенный дескриптор, представляющий собой слово или символ, по которому можно осуществить поиск документальной информации. Первые информационно-поисковые тезаурусы представляли собой печатные издания, например, в качестве информационно поискового тезауруса могут выступать различные справочники, в последнее время наблюдается колоссальный рост автоматизированных информационно-поисковых тезаурусов на электронных носителях, в которых поиск нужной информации осуществляется автоматически при помощи соответствующих программ. Примером последних выступают различные электронные энциклопедии и словари, а также информационно-поисковые системы, работающие on-line.

Понимание тезауруса как системы знаний о действительности обострило проблему изучения смысловой информации, не перестававшую быть актуальной на протяжении XX в., и роли естественного языка как носителя информации о мире. Понятие «тезаурус» приобрело междисциплинарный характер и стало трактоваться весьма широко как система знаний о действительности, имеющаяся у человека (индивидуального носителя информации) или группы людей.

Поскольку в основе всех рассмотренных представлений о сущности тезауруса лежит система языка, это позволило сформулировать предположение, что тезаурус – это объективно существующая языковая данность, функционирующая как информационная система в человеческой среде, определяющая структуры знания отдельного индивида и общества в целом на определенном этапе его развития, приблизительной моделью которой, очевидно, является словарь.

Поскольку в лингвистике слово «тезаурус» в первую очередь ассоциируется с особым типом словарей, правомерно начать рассмотрение истории развития понятия «тезаурус» с истории тезаурусных словарей. Поскольку тезаурус определяется через понятие «идеографический словарь», будем использовать слова «тезаурус» и «идеографический словарь» как синонимы.

Подводя итог краткому обзору идеографических словарей прошлого, выделим основные идеи, повлиявшие на современные представления о сущности тезауруса:

1. Тезаурусы представляют собой первые попытки осмысления и систематического представления лексики языка. Лексическая система предстает в них как базирующаяся на строгих логических отношениях, преимущественно родовидовых, позволяющих выстраивать слова языка в иерархию соподчиненных классов. Иерархия родовидовых отношений трактуется как основная структура человеческого мышления. Показательно, что первыми словарями человеческих языков являются именно словари со строгой иерархией классов слов; родовидовые отношения впоследствии были положены в основу дефиниций, предлагаемых в толковых словарях. Итак, система лексики языка – это иерархическая система понятий и соответствующих слов.

2. В ходе истории создания тезаурусных словарей сформировались три основных типа тезаурусов: 1) тематические словари, в которых представлены слова, объединенные некоей общей темой; 2) собственно тезаурусы, или идеографические словари, отражающие ход мысли от общих мыслительных категорий и понятий к словам языка; 3) аналогические словари, совмещающие алфавитный принцип изложения слов и разбиение слов по смысловым группам.

Данные типы, по сути, отражают ход исследовательской мысли по осознанию устройства лексики вообще: сначала от частного к общему (от объединения конкретных слов в темы и дальнейшего осмысления стоящих за темами философских категорий), а затем вновь, к отдельному слову как центральной единице, от которой следует отталкиваться при поиске других слов.

Способы представления слов в тезаурусах (разбиение на классы с соответствующим буквенным или цифровым индексированием) предопределило развитие идеи изучения языка в формальных категориях логики, идеи записывания выражений языка на формальном языке логики и идеи создания искусственных формальных языков, которые в последствии привели к появлению компьютерных языков и попыткам конструирования систем искусственного интеллекта.

Тезаурус – это, прежде всего, система. Система чего? – Возможны варианты: система лексики, эксплицирующая только словесные связи; система понятий, отражающая состояние знания (в том числе научного) на определенном этапе; система знания, имеющаяся у человека и обнаруживаемая путем психолингвистического эксперимента.

Соответственно, имеется три типа современных тезаурусных словарей:

1) словари, раскрывающие связи от слова к слову (тезаурусы английского языка, кроме словаря Роже, а также словари синонимов, антонимов, сочетаемости);

2) словари, раскрывающие связи от понятия к слову (словарь Роже, идеографический словарь О. С. Баранова, тематический словарь, электронные тезаурусы типа WordNet);

3) словари, раскрывающие связи от слова-стимула к слову-реакции (Edinburgh Associative Thesaurus, Русский ассоциативный словарь). По сути, за этими вариантами понимания тезауруса встает семиотическое соотношение «знак – объективное понятийное содержание – субъективное ментальное содержание».

Можно предположить, что печатные идеографические словари и электронные тезаурусы представляют собой «внешние», «объективированные» тезаурусы, в которых представлены лексические отношения, изученные исследователями и являющиеся результатом их сознательной умственной деятельности по упорядочиванию лексической информации языка. Ассоциативные словари проливают свет на устройство «внутреннего», «субъективного» тезауруса человека, существующего, скорее, стихийно. Первые строятся преимущественно на основе иерархических отношений, вторые более ярко фиксируют синтагматику словесных отношений. Думается, что собственно языковой тезаурус должен как-то совмещать обе структуры.

Итак, можно сформулировать предварительное положение, что тезаурус – это семиотическая система, элементы которой – словесные знаки. Соответственно, по аналогии с фундаментальными вопросами семиотики о соотношении знака, реальности и ментальности, встают вопросы: как существует реальность в тезаурусе (или тезаурус в реальности?) и как существует тезаурус в голове человека?

Тезаурусы, созданные на основе логических онтологий, наводят на мысль, что объективное существование системы понятий – само собой разумеющееся. Однако несовпадение выводимых в таких словарях понятийных систем заставляет усомниться в правильности такого решения. Тезаурусы, идущие от слова к слову, предполагают феноменологическое основание: невозможно восстановить объективную систему понятий, поэтому восстанавливается только то, что реально дано наблюдателю – словесные связи. Данный подход более обоснован, однако, несовпадение перечня связей для отдельного слова в разных словарях свидетельствует о значимости фигуры исследователя-составителя и о важности его субъективного восприятия релевантности связей, его субъективной оценки связей как «стоящих/нестоящих включения в словарь». Наконец, тезаурус, раскрывающий словесные реакции испытуемых в ходе психолингвистического эксперимента еще более абсолютизирует субъективность видения словесных связей, так как предлагаются словесные реакции конкретных людей. Создатели словаря преодолевают проблему субъективизма полученных данных частотным анализом словесных реакций, но вопрос о соотношении реальности и тезауруса остается открыт.

Вопрос о том, как существует тезаурус «в голове» человека, не имеет четкого ответа, если опираться только на данные проанализированных словарей, однако, можно выделить несколько «подсказок», в каком направлении стоит двигаться, изучая данную проблему.





Во-первых, все тезаурусные словари (как и словари вообще) служат инструментом поиска информации. Однако словарь сам ничего не ищет – ищет пользователь словаря по заданному алгоритму. Если поиск затрудняется

(а это один из главных недостатков печатных идеографических словарей), значит, алгоритм поиска, предложенный в словаре, не совпадает с алгоритмом поиска, из которого исходит пользователь. Электронные тезаурусы, напротив, ищут слово сами и, на первый взгляд, в состоянии предложить решение проблемы быстрого поиска нужного слова. Однако это, скорее, следствие качественной инженерной разработки языков программирования, нежели результат нахождения языкового механизма. Пользователь словаря не видит логическую цепочку понятий, выстраивающих путь к искомому слову, и потому нельзя судить, совпадает ли предлагаемый программой алгоритм поиска с внутренним алгоритмом языкового тезауруса или с ментальным алгоритмом свойственным системе знания человека. Каков этот алгоритм – пока неясно. Можно усмотреть рациональное зерно в обращении к алгоритму поиска по алфавиту, поскольку печатными тезаурусами аналогического типа пользоваться легче, чем собственно идеографическими словарями. При составлении алфавитного указателя автор словаря исходит только из материальной оболочки слова, точно также и компьютерная программа исходит только из материальной составляющей знака, подыскивая нужное пользователю слово. Можно предположить, что и ментальный алгоритм поиска слов связан с материальной оболочкой слов, а не только со значением, как принято думать.

Во-вторых, поиск информации всегда связан с категорией памяти, не случайно авторы тезаурусных словарей пишут, что предлагаемые издания необходимы для извлечения нужного слова из памяти. Что собой представляет «память на слова», тоже, пока сказать трудно, но, очевидно, она базируется на связях слов, поскольку тезаурусные словари предлагают помощь в извлечении слов из памяти путем представления связей слов с другими словами. По этому поводу стоит вспомнить две метафоры, при помощи которых составители тезаурусов именуют эти связи как совокупную целостность: «сеть» и «процессор». Думается, эти метафоры весьма эффективны для дальнейшего осмысления идеи памяти. Также, основываясь на данных ассоциативных тезаурусов, можно предположить, что «память на слова» формируется в меньшей степени из системы понятий, и в большей степени из заученных в «готовом виде» высказываний. Поэтому весьма перспективным является приведение в тезаурусах исчерпывающего перечня кол локаций, сочетаний, устойчивых выражений, валентностей, прецедентных текстов и других «готовых к употреблению» объединений слов. Чтобы понять, как все многообразие языковых выражений входит в человеческую память, необходимо найти общее во всех этих разных с лингвистической точки зрения объединениях слов.

Возвращаясь к представлению о тезаурусе как о системе, встает еще один важный вопрос – об элементах и связях этой системы. Что касается элементов, то, предварительно можно ответить, что это слова, но слова, растворенные в своих связях, причем не только словесных, но и понятийных. Понятые так слова вполне уместно называть кол локациями, так как слова в системе тезауруса не существуют как само ценности, их ценность определяется количеством и качеством их связей. В данном

случае составители словарей обращаются к частотности, как формальному показателю стабильности связей, на который можно опереться и при отборе заглавных слов, и при отборе слов, составляющих содержание статьи. Кол локациями, достойными отображения в тезаурусе, признаются те, которые соответствуют критерию стабильности в силу высокой частоты их возникновения.

Вопрос о качестве тезаурусных связей – наиболее трудный. Анализ словарей показывает, что это могут быть связи собственно словесные и понятийные. По поводу того, какие из них главные, а какие вспомогательные, можно ответить двояко. Если изначально встать на позицию признания существования объективных понятийных связей, то словесные связи – только средство их выражения. Однако, структура представления понятийных связей в идеографических словарях наводит на мысль об отсутствии четкой логики этих связей, которая неизбежно предполагается самим фактом их объективности. Соответственно, возникает второй вариант ответа: возможно, иерархия понятийных связей не существует объективно, а мотивирована словесными связями языка, преломленными через их субъективную оценку составителем словаря (главным образом на основе критерия частотности).

Что касается собственно словесных связей, то, исходя из содержания проанализированных тезаурусных словарей, можно сказать, что в сети тезауруса слова могут быть объединены любыми известными лексическими связями (парадигматическими и синтагматическими), а также ассоциативными связями. Причем трудно сказать, какие из них более важны, поскольку, в одних словарях раскрываются преимущественно синонимические связи, в других – преимущественно контекстуальные, в-третьих синонимы определяются по контексту, в-четвертых контекст иллюстрирует использование определенного синонима и т. Д.

Итак, тезаурус, по предварительному представлению – это система словесных связей, семиотическая по природе, со всеми вытекающими отсюда последствиями.

Наиболее подходящими методами исследования при изучении онтологической сущности тезауруса являются эмпирические (сбор и обработка языковых фактов), а при построении модели – теоретическое обобщение и логическое обоснование выводов.

Думается, что для большей эффективности обоих направлений исследования, а также с целью преодоления обозначенных проблем в понимании сущности тезауруса, стоит обратиться к общенаучному учению о системе. Разногласия в принципах построения словарей проистекают из того, что в лингвистике имеется несколько «статическое» представление о лексической системе, сводящееся к перечислению парадигматических и синтагматических связей слов.

Для того, чтобы более наглядно представить целостность словесной системы и более глубоко осмыслить принципы ее жизнедеятельности, необходимо включить лингвистическое знание о системе лексики в более широкий научный контекст изучения систем. Для этого стоит обратиться к общей теории систем, тем более, что именно с развитием общей теории систем связано комплексное изучение информации, становление теории информации и информатики, которые в середине 20 века стали научной базой для конструирования информационно поисковых тезаурусов, а также позволили сформировать представление о тезаурусе как о семантической

информационной системе, содержащей метаинформацию о функционировании самой информации.

### References:

1. Фрумкина Р. М. Психоллингвистика: Учеб. для высш. учеб. заведений.
2. Фрумкина Р.М. – М.: Издательский центр «Академия», 2001. – 230 с.
3. Фрумкина Р.М. Есть ли у лингвистики своя эпистемология // Р.М. Фрумкина // Язык и наука конца XX века. – М.: Институт языкознания РАН, 1995. – С. 74-117.
4. Шайкевич, А.Я. Тезаурус // Языкознание: Большой энциклопедический словарь. – М.: Большая Российская энциклопедия, 1998. – С. 506- 507.
5. Шрейдер, Ю.А. Системы и модели // Ю.А. Шрейдер, А.А. Шаров. –М.: Радио и связь, 1982. – 152 с.
6. Шрейдер, Ю.А. Ценности, которые мы выбираем. Смысл и предпосылки ценностного выбора // – М.: Эдиториал УРСС, 1999. – 208 с.
7. Баранов, А.Г. Функционально-прагматическая концепция текста.
8. А.Г. Баранов. – Ростов-на-Дону: Изд-во Ростов. ун-та, 1993. – 180 с.