

## **ПЕРСПЕКТИВЫ РАЗВИТИЯ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ТЕХНОЛОГИЙ**

**Ю.И. КИСЛЕНКО**

Рассматривается класс информационных технологий обработки естественно-языковой информации. В общем случае подобные системы моделируют речевое поведение человека, и их информационное ядро должно представляться совокупностью лингвистического процессора и базы знаний, работающих друг на друга. С позиций предложенной автором формальной модели языка анализируются перспективы информационных технологий указанного класса.

### **ВВЕДЕНИЕ**

Повышенный интерес к формированию информационных технологий обработки естественного языка (ЕЯ) и достаточно скромные успехи на этом поприще обуславливают интеграцию усилий специалистов различных направлений (информационщиков, лингвистов, психологов, физиологов и др.) на пути к познанию феномена языка и моделирования речевой деятельности. Такой интеграционный подход как по составу организаторов, так и по тематике докладов, характерен для Первого семинара по компьютеризации естественных языков (КЕЯ), состоявшегося в Варне в сентябре 1999 г. Организатором этого семинара был пионер болгарской компьютерной школы профессор Шишков Димитар Петров.

На семинаре [1] были сформулированы лишь общие направления компьютерной обработки ЕЯ-информации. Однако уже тогда наметились приоритеты этой важной и сложной области. Это — вопросы, связанные с фундаментальными проблемами организации естественного языка на всех уровнях (структурном, морфологическом, семантическом), сохранения письменных памятников, извлечения знаний из текстов с последующим их представлением в базах данных и знаний, синтеза и анализа речевой информации, задачи нормирования языков, интегрирования разноплановой информации и т.п. Конечно же, можно заметить значительную неопределенность в выборе задач, отсутствие системного подхода, несколько неоправданную эйфорию в надежде на незамедлительные результаты. Вполне объяснимая ситуация для начального этапа формирования программы исследований, особенно в такой сложной области, как моделирование речевого поведения человека.

Цель настоящей работы — представить современное состояние класса информационных технологий, обработки ЕЯ, показать, что в данной области моделирования интеллектуальной деятельности человека намечаются обнадеживающие перспективы, позволяющие достаточно уверенно прогнозировать направления исследований и конструктивного решения некоторых фундаментальных вопросов обработки ЕЯ-информации. Речь идет о формальной модели языка, разработанной автором в НТУУ «КПИ» и представленной на Первом семинаре в качестве продуктивной основы создания современных ЕЯ-технологий. За эти годы модель приобрела более законченный вид, получены новые результаты в формировании ЕЯ-технологий.

### ТЕХНОЛОГИИ ОБРАБОТКИ ЕЯ-ИНФОРМАЦИИ

Среди современных информационных технологий особый класс составляют системы обработки естественно-языковой информации. К этому классу принадлежат системы: *информационного поиска в INTERNET; экспертные; автоматического перевода; информационно-поисковые; обработки (синтез/анализ) текстовой информации; обработки (синтез/анализ) речевой информации; «понимания» речевого сообщения; интегрирующие в себе обработку образной и символьной информации.* Перечень подобных систем завершается *естественно-языковым интерфейсом и системой накопления знаний*, что является приоритетным направлением в создании компьютеров пятого поколения, которые, к сожалению, так и не реализованы до сих пор.

Именно последние два направления представляют наибольший интерес в создании ЕЯ-технологий, ибо концентрируют в себе все узловые проблемы предыдущих: первое из них — не что иное, как система обработки ЕЯ-информации (лингвистический процессор (ЛП)), второе — база знаний (БЗ), в которой формируется модель внешнего мира, а вместе они составляют *информационное ядро* технологий обработки ЕЯ.

Все перечисленные системы так или иначе связаны с моделированием речевого поведения человека — одной из наиболее сложных форм его интеллектуальной деятельности. *Речевая деятельность в общем случае базируется на двух составляющих, определяющих как наши знания об организации языка (условно эту составляющую определим как языковую компетенцию, ответственную в дальнейшем за формирование ЛП), так и знания относительно внешнего мира (представленные в БЗ в виде соответствующей модели), и любые проявления речевой деятельности (будь-то синтез или анализ речевого сообщения) обязательно актуализируются лишь при условии диалектического взаимодействия обеих составляющих — ЛП и БЗ.* Следовательно, и информационное обеспечение соответствующих технологий должно базироваться на этих составляющих и их взаимодействии.

### ТЕКУЩЕЕ СОСТОЯНИЕ ЕЯ-ТЕХНОЛОГИЙ

Появление мощной вычислительной техники снова (в который раз!) возрождает иллюзии быстрого и эффективного моделирования интеллектуальных возможностей человека, в частности, моделирования речевой деятельности, что напрямую связано с разработкой ЕЯ-технологий. В западном мире в по-

следнее время наблюдается очередной всплеск эмоций и надежд, связанных с моделированием сложных интеллектуальных, эмоциональных и психических функций человека. Формируется целое поколение «интеллектуальных агентов», ответственных за выполнение отдельных узкоориентированных операций и процедур, «плодятся» информационные роботы, которые размножаются в сети INTERNET и поступают с информацией по своему усмотрению, формируются отдельные домены знаний в рамках общей онтологии и т.п.

Однако многолетний опыт использования вычислительной техники как средства моделирования убеждает, что эффективные модели строятся там, где четко и однозначно представлена информационно-логическая модель исследуемого объекта. Без внутренней стройности и логической завершенности таких моделей не может быть и речи об эффективном создании соответствующих технологий. А в нашем случае объектом исследования (следовательно, и моделирования) выступает речевая деятельность человека. Совокупные знания об этой сфере пока совершенно неадекватно отображают потенциальные возможности речевой деятельности, и тем более результаты классической лингвистики весьма далеки от того, чтобы их можно было переложить на строгий язык компьютера. Это позволяет с определенной долей здравого скепсиса относиться ко всякого рода «прожектам» в области моделирования речевой деятельности. Здесь еще очень много вопросов. Эффективных технологий в этой области не будет, пока не появятся достаточно строгие модели речевой деятельности, опирающиеся на основные информационные составляющие человека (знания о языке и мире), обладающие достаточной объяснительной способностью и сводящие все накопленные знания (не только в лингвистике, но и в пограничных областях) в четкую непротиворечивую систему. Только в случае выполнения этих требований можно предметно говорить о возможности формирования соответствующих технологий.

Реальное же положение дел в области ЕЯ-технологий таково: практически все разработки ориентированы на человеко-машинный вариант работы (редакторы, переводчики, системы поиска и т.п. требуют участия человека на завершающих этапах). Это следует из того, что, во-первых, все современные технологии обработки естественного языка уверенно работают лишь с идеальным языковым материалом [2], во-вторых, классическая лингвистика в создание компьютерных технологий внесла довольно-таки незначительный вклад [3]. В концентрированном виде оценка текущего состояния классической лингвистики в области семантических исследований как «кризисного» описана в работе [4].

## **ПРИЧИНЫ НЕУДОВЛЕТВОРИТЕЛЬНОГО СОСТОЯНИЯ ЕЯ-ТЕХНОЛОГИЙ**

Речевая деятельность представляется одной из наиболее сложных форм интеллектуальной деятельности человека, которая в символической форме отображает процессы восприятия, переработки и интеграции разноплановой информации. Речевая деятельность [5] в качестве равноправных компонен-

тов содержит процессы как порождения (синтеза), так и восприятия (анализа) речевой информации. *Актуализация речевой деятельности возможна лишь при условии эффективного взаимодействия важнейших информационных составляющих, представляющих как знания об устройстве языка и законах его функционирования (ЛП в нашем представлении), так и знания о внешнем мире (модель мира в БЗ).* Если мы хотим моделировать речевую деятельность человека хотя бы в каком-то приближении, предварительно необходимо построить модели каждой из этих составляющих и обеспечить их диалектическое взаимодействие. Последнее означает, что ЛП обеспечивает обработку входной ЕЯ-информации, используя накопленные знания в БЗ, а БЗ, в свою очередь, ответственна за накопление знаний и способствует работе ЛП. Лишь при наличии обеих составляющих можно корректно ставить вопрос о моделировании речевой деятельности человека в достаточно полном объеме [6]. Пока же все существующие разработки в области ЕЯ-технологий учитывают и моделируют лишь первую составляющую — наши знания о языке (и то не лучшим образом), «забывая» совершенно о другой неотъемлемой составляющей — знаниях о внешнем мире, которые обеспечивают надлежащую интерпретацию текстовой информации.

Неудовлетворительное состояние дел в области ЕЯ-технологий, конечно же, в первую очередь определяется *сложностью объекта исследования* — речевого поведения человека, обеспечивающего как порождение текста, так и его понимание (это процессы одного порядка сложности). Причины такого состояния современных ЕЯ-технологий кроются в следующем:

- *отсутствие эффективной методологии* исследования речевой деятельности человека;
- *неадекватность наших знаний о языке реалиям речевой организации;*
- *отсутствие четкой системы* в организации наших знаний об устройстве языка и особенностях его функционирования;
- *неполнота информационного обеспечения* современных ЕЯ-технологий, не учитывающего диалектическое взаимодействие важнейших составляющих информационного обеспечения речевой деятельности — знаний о языке и внешнем мире.

## ПУТИ ВЫХОДА ИЗ КРИЗИСНОЙ СИТУАЦИИ

Четкое осознание причин неудовлетворительного состояния дел в области моделирования речевого поведения, учет критических замечаний и пожеланий признанных авторитетов открывают новые направления в исследовании речевой деятельности и позволяют внести конструктивные предложения по созданию эффективных ЕЯ-технологий.

Существующие знания о языке базируются на результатах *классической лингвистики*, которая отличается значительной самодостаточностью и отстаивает принцип: «язык — это высшая форма интеллектуальной деятельности человека, и через язык мы познаем самого человека» [3]. Данный подход практически исповедует идеологию «черного ящика» в классиче-

ском его варианте, когда наблюдателю представляется лишь его выход — текстовая информация, а что происходит внутри ящика (в человеческом мозгу) и что подается на вход такой системы совершенно не интересует исследователей. Такая концепция утверждает: наблюдая речевой материал (как конечный продукт речевой деятельности), мы сможем понять все тонкости владения языком и его организацию. Эта точка зрения вызывает определенные возражения из-за своей ограниченности и замкнутости. Кроме того, данный подход исключает из рассмотрения анализ процессов порождения текста.

В настоящее время все большую популярность и конструктивность приобретает *информационный подход* к анализу речевой деятельности, учитывающий «человеческий фактор» и утверждающий, что только тщательно изучив все системы восприятия и обработки произвольной информации как в отдельности, так и во взаимосвязи, и проанализировав, по возможности, собственно этапы синтеза и анализа речевого сообщения, мы сможем определить весьма сложное интеграционное явление, коим является речевая деятельность. Этот подход несколько смягчает идеологию «черного ящика», и чем более «прозрачным» становится он для нас с информационной точки зрения (т.е. чем больше мы будем знать о функциях зрительного, акустического и других анализаторов), тем полнее будут наши знания об общей организации и функционировании языка. По сути, данный подход учитывает информационные процессы в нейронных структурах коры головного мозга, указывая как на конкретные особенности систем восприятия и обработки информации, так и на интеграционный характер их взаимодействия.

Вероятно, выход из тупиковой ситуации кроется в симбиозе классической лингвистики и информационного подхода. Это означает, что все накопленные знания об устройстве языка необходимо пропустить через *информационную призму речевой деятельности*. При этом, хаотическое нагромождение множества фактов превращается в стройную систему с удивительным порядком. Многие спорные вопросы приобретают ясность и совершенно иное толкование. Конечно, ради такого сотрудничества необходимо отказаться от снобизма классической лингвистики и с пониманием отнестись к усилиям специалистов различных пограничных областей, пытающихся со своей стороны понять такое архисложное явление, как речевая деятельность. Этот синтетический подход только начинает формироваться, однако конструктивные предпосылки его появления уже налицо и они встречаются в различных областях исследований, что вызвано критикой отдельных направлений нашей науки о языке и поиском путей выхода из создавшегося положения.

В таком аспекте и следует рассматривать предложения ведущих специалистов классической лингвистики и информационного подхода относительно оценки текущего положения науки о языке и перспектив выхода из кризисного состояния. На этом пути следует указать важнейшие вехи оценки речевой деятельности специалистами различных направлений.

1. Л.И. Астахова [7] практически первой забила тревогу относительно текущего положения дел в области синтаксиса и представила в концентрированном виде *критическую оценку современного состояния синтаксиса простого предложения*, указав при этом, что общественное мнение уже дав-

но подготовлено к отрицанию установившегося деления предложений на простые и сложные. Вывод таков: *необходимо четко определить сам объект синтаксических исследований и заменить основания существующего подхода.*

2. Н.М. Перцов [3] убедительно критикует современные семантические концепции и дает реалистическую оценку устремлений и достижений классической лингвистики, подчеркивая ее изоляционизм и весьма скромный вклад в область прикладных разработок. *Путь преодоления кризисного состояния — в интеграционном подходе к учету достижений специалистов, работающих в различных пограничных областях исследования речевой деятельности.*

3. В.А. Звезгинцев [8] дает глубокий анализ существующих подходов к рассмотрению структурного уровня предложения, вскрывает их недостатки и показывает, что решение важнейших вопросов лингвистики и выход из кризисного состояния кроется в учете триединства: *мышление, язык, действительность.* Его «*пресуппозиции*» выходят за уровень синтаксиса предложения и намечают (весьма призрачную) взаимосвязь между базой знаний и лингвистическим процессором как системой обработки языкового материала.

4. Е.С. Кубрякова [9] последовательно отстаивает необходимость учета «*человеческого фактора*» в языке, т.е. выходит за пределы традиционного объекта исследования классической лингвистики, пытается каким-то образом сломать стереотип «*черного ящика*», который сложился в классической лингвистике, и перспективы видит в *тщательном исследовании речевой деятельности* человека, содержащей в качестве равноправных составляющих синтез и анализ языкового материала.

5. Ю.Д. Апресян [10], хотя и является идеологом направления семантических поисков, которое критикует Н.М. Перцов, приоткрывает тем не менее путь к учету сенсорной информации и ее интеграции с символьной, вводя в орбиту лингвистических интересов «*изобразительные средства семантики*». Это послужило продуктивным началом построения автором модели «*Действительность — Текст*» для узкой предметной области пространственных отношений и открыло интересные перспективы в области семантических исследований.

Заслуживают серьезного внимания и предложения сторонников информационного подхода к дальнейшим перспективам формирования ЕЯ-технологий.

6. Г.П. Мельников [6] перенес в практическую плоскость исследование соотношения *мышление — язык — действительность*, дав весьма продуктивную модель речевой деятельности. Весьма своевременной и продуктивной в этой модели представляется концепция *взаимодействия лингвистического процессора, учитывающего языковую компетенцию с базой знаний, где хранится модель внешнего окружения.*

7. А.В. Анисимов [11] перспективы прикладной лингвистики видит в структурном объединении всех важнейших разделов лингвистики (*морфологии, синтаксиса и семантики*) с целью повышения эффективности синтеза и анализа текстовой информации. Продуктивной считает концепцию

базы знаний в виде *онтологическо-семантической сети*. Весьма значительный интерес представляют также исследования в области *рекурсивной организации текстовой информации* [12].

8. В.П. Гладун [13] весьма последовательно и продуктивно работает в области моделирования нейронной организации баз знаний для представления символической информации. Очень плодотворной может оказаться концепция *растущих пирамидальных структур* для моделирования процесса накопления знаний.

9. Г.М. Зенкин и А.П. Петров [14] опубликовали цикл работ, посвященных вопросам *интеграции образной и символической информации*. В этом цикле профессионально проанализирован информационный аспект работы зрительного анализатора в задачах описания внешнего мира. Можно также выделить раздел, посвященный *«геиштальт-синтезу»*, где интегрируются особенности обработки зрительной и символической информации. К этому же пограничному направлению следует отнести работу В.Д. Глезера [15], где впервые предложена информационная модель процедуры синтеза описания отдельной ситуации внешнего мира.

Практически все указанные моменты критического и перспективных планов созвучны позиции автора относительно организации речевой деятельности и тем или иным образом учтены в разрабатываемой им формальной модели. Отдавая приоритет информационному подходу, конечно же, нельзя сбрасывать со счетов весь накопленный (как положительный, так и отрицательный) опыт классической лингвистики.

Важнейшей платформой предложенного подхода являются следующие аксиомы:

- *системная организация языка определяется системной организацией памяти человека;*
- *язык в общем случае является средством отображения как внешнего мира, так и информационного состояния нашего мозга, что налагает определенные ограничения на структурный уровень его организации;*
- *структурная организация текстовой информации во многом определяется информационными возможностями систем восприятия внешнего мира (в частности, особенностями организации сенсорного уровня восприятия информации);*
- *языковой материал в общем случае отличается рекурсивной схемой организации, где элементом рекурсии выступает четко определенная структура.*

Это, по сути, отдельные элементы аксиоматики речевой деятельности, представленные еще на Первом семинаре, весьма важные для формирования ЕЯ-технологий. Они стали плодотворной платформой для создания формальной модели языка [16].

Данный подход снимает ряд существенных противоречий современных грамматик, с единых позиций предлагает конструктивные решения в области создания лингвистического процессора и БЗ, может служить продуктивной основой формирования современных технологий, ориентированных на обработку ЕЯ-информации.

## ФОРМАЛЬНАЯ МОДЕЛЬ КАК МЕТОДОЛОГИЧЕСКАЯ ОСНОВА ЕЯ-ТЕХНОЛОГИЙ

На кафедре технической кибернетики факультета информатики НТУУ «КПИ» автором разработана формальная модель, которая с системных позиций анализирует такое сложное явление, как язык, четко определяет конструктивные элементы и их взаимосвязь и обладает достаточной объяснительной силой. Эта модель отличается от других следующими ключевыми позициями.

1. Предлагаемая модель базируется на общих принципах восприятия и обработки разноплановой информации человеком (зрительной, акустической, тактильной, символьной) с последующей ее интеграцией.

2. Данный подход определяет модель филогенеза языка, объективной основой которой являются важнейшие этапы онтогенеза речи ребенка, тогда как все существующие модели носят сугубо субъективный характер.

3. Ключевым понятием модели представляется некий *синтаксический шаблон* — базовая семантико-синтаксическая структура, определенная на формальном уровне и выступающая в качестве основы формирования произвольного сообщения. *Базовая семантико-синтаксическая структура* — это двусоставная монопредикатная структура описания произвольной ситуации внешнего мира, все элементы которой не выходят за атрибутивный уровень их описания. Такое описание отличается от известных (а их к настоящему времени насчитывается около трехсот) двумя позициями: *во-первых, это максимально полная схема представления произвольной ситуации, во-вторых, элементами подобной структуры не могут быть словосочетания.*

4. Модель рассматривает, с одной стороны, возможные схемы трансформирования базовой структуры, которые определяют монопредикатный уровень формирования сообщения, а с другой — определяет потенциально возможные схемы их взаимосвязи (полипредикатный срез формирования сообщения). *Дихотомия монопредикатный/полипредикатный уровни значительно полнее и четче существующего разграничения простое/сложное предложение).*

5. Предлагаемый подход рассматривает связность текстовой информации как следствие отображения целостности внешнего мира.

6. Модель снимает проблему словосочетания, которое рассматривается как один из возможных вариантов структур монопредикатного уровня. Это одна из сложнейших проблем современного языкознания, которая *была сформулирована еще в 1660 г., но не разрешена и по сей день.*

7. Предлагаемая модель определяет произвольный текстовый материал как рекурсивно-организованную систему, где *элементом рекурсии выступает базовая структура*, что позволяет выйти на формальный уровень представления структуры сообщения, где центральным становится единый синтаксический шаблон [17].

8. Модель определяет аксиоматику формирования структурного уровня организации текста, основанную на признании факта, что язык — универсальная форма представления произвольной информации относительно

окружающего мира. Возможность построения аксиоматики переводит языкознание из разряда нечетких гуманитарных дисциплин в разряд точных наук.

9. Предлагаемый подход отличается от существующих четко выраженной системной организацией представления структурного уровня сообщения, базирующегося на иерархии: базовая структура, монопредикатный уровень, полипредикатный уровень.

*Каждый из этих пунктов — практически, важнейшая проблема, сформулированная, но нерешенная классической лингвистикой на современном этапе*, и одновременно — конструктивные элементы информационных технологий, ибо формальная модель собственно и создавалась, в первую очередь, для информационных технологий. По широте охвата тематики предложенная модель выходит за рамки представления структурного уровня организации текстовой информации, рассматривая многие важнейшие вопросы актуализации речевой деятельности человека. Таким образом, предложенная модель представляется продуктивной методологической основой исследования речевой деятельности человека и, следовательно, основой формирования современных ЕЯ-технологий.

## **ИСПОЛЬЗОВАНИЕ ФОРМАЛЬНОЙ МОДЕЛИ В ЕЯ-ТЕХНОЛОГИЯХ**

Следствиями предложенного подхода к структурной организации языкового материала явились весьма конструктивные предложения.

1. Модель, практически, определяет возможность формирования структуры ЛП, учитывающего формализмы предложенного подхода (базовая структура, монопредикатный, полипредикатный уровни) и ориентированного, главным образом, на обработку базовой структуры.

2. Указанный подход определяет также *архитектуру БЗ, где единицей восприятия, накопления и обработки информации выступает базовая семантико-синтаксическая структура*. В БЗ закладывается концепция структурного единства морфологии, синтаксиса и семантики и возможности взаимодействия с ЛП.

3. Одним из важнейших прикладных аспектов является возможность приведения произвольного текста к *канонической форме*, представленной совокупностью базовых структур или их трансформаций. Данное преобразование получается в результате декомпозиции входного текста по базовым семантико-синтаксическим структурам с учетом диалектического взаимодействия ЛП и БЗ. Важность подобного преобразования определяется тем, что для флективных языков исключается жесткий порядок актуализации синтаксического шаблона. Это автоматически приводит к «проклятию многомерности», как только мы пытаемся перечислить возможные варианты заполнения такого синтаксического шаблона. Для структуры шаблона из восьми составляющих возможные варианты актуализации сообщения достигают астрономических цифр в сорок миллиардов вариантов [18]. Вот почему при разработке ЕЯ-технологий важно уметь *приводить произвольную структуру к ее единственному варианту*. Такая задача входит в компетенцию ЛП. В общем случае *ЛП должен решать задачу декомпозиции сообще-*

ния по базовым структурам с последующей процедурой их канонизации. Каноническая форма однозначно определяет порядок учета составляющих синтаксического шаблона.

4. Важным следствием предложенного подхода явилась возможность использования *изобразительных средств семантики* для моделирования соотношения «Действительность — Текст» в том или ином направлении (здесь моделируется связь между образной и символической БЗ).

## ПЕРСПЕКТИВЫ РАЗВИТИЯ ЕЯ-ТЕХНОЛОГИЙ

Использование предложенной модели в качестве методологической основы анализа речевой деятельности позволяет совершенно по-иному с конструктивных позиций подойти к формированию информационного ядра (ЛП — БЗ) ЕЯ-технологий, что дает возможность по-новому планировать развитие ЕЯ-технологий во многих приложениях.

1. **Разработка лингвистического процессора** с позиций предложенной модели базируется на учете формализмов определения базовой структуры, монопредикатного и полипредикатного уровней. Использование данной модели позволяет процедуру анализа текстовой информации перевести из области случайного поиска нечетко определенных структур в разряд прогнозируемого поиска конечного перечня формально определенных структур.

2. **Формирование базы знаний.** Предложенная модель позволяет совершенно с других позиций подойти к формированию архитектуры БЗ, где единицей восприятия, накопления и обработки информации выступает базовая семантико-синтаксическая структура, а собственно знание представляется последовательностью однотипных базовых структур. Разработка ЛП и БЗ должна формироваться с учетом их диалектического взаимодействия. Связка ЛП–БЗ, по существу, основа формирования ЕЯ-интерфейса и системы накопления знаний.

3. **Каноническое представление текста** — один из важнейших этапов практически любых информационных технологий обработки ЕЯ-информации. Результат преобразования — декомпозиция входного текста по базовым структурам с указанием их связи. Причем каждая из таких структур «разворачивается» в строго определенном порядке учета как атрибутивных членов, так и актантов совместно с сирконстантами. Такая процедура «канонизации» произвольного сообщения позволяет последнее привести к единственно возможному варианту из практически неисчислимого количества схем актуализации описания произвольной ситуации. По сути, данная процедура представляется обязательным этапом предварительной обработки текстов любых ЕЯ-технологий. Можно предположить, что процедура канонизации входного сообщения автоматически актуализируется в нейронных структурах нашего мозга, иначе как объяснить столь эффективные процедуры обработки текстовой информации человеком.

4. **Поиск информации в INTERNET.** Технология поиска — это последовательность этапов: стандартная процедура поиска по ключевым словам или их комбинации; после нахождения релевантных документов выпол-

няется процедура канонизации входного текста и найденных документов и, наконец, процедура информационного поиска канонизированного запроса на массиве канонизированных текстов. Принципиальное отличие такой схемы от существующих процедур в том, что поиск по ключевым словам заменяется поиском среди канонизированных структур.

**5. Информационно-поисковые системы.** Технология поиска аналогична предыдущему случаю, где массивы документов и запросы предварительно подвергаются процедуре канонизации.

**6. Автоматический перевод** состоит из последовательных этапов: канонизация входного текста на Я1, перевод отдельных элементов канонического представления с Я1 на Я2, переход к каноническому представлению текста на Я2. В отличие от поисковых процедур, где технология скрыта от пользователя, в задачах перевода после выполнения преобразований над входным текстом на Я1, необходимо уже на языке Я2 снова вернуться от канонизированного представления к актуализации сообщения за схемой, представленной входным текстом.

**7. Системы обработки речевой информации.** Проблемой подобных технологий является адекватность анализа сложных конструкций. Каноническая форма представления текста используется для более качественного анализа входного текста. Кроме того, в системах обработки речевой информации (как для синтезаторов, так и анализаторов) явно ощущается отсутствие второй составляющей речевой деятельности — БЗ.

**8. Интеграция систем обработки речевой и текстовой информации (системы распознавания речи).** Предлагается использовать связку ЛП — БЗ в системах обработки речевой информации (БЗ в таких системах практически не используются).

**9. Интеграция систем обработки зрительной и символьной информации.** Продуктивной представляется концепция использования изобразительных средств в качестве создания семантического описания отдельных предметных областей. В рамках предложенной модели проверена возможность реализации соотношения «Действительность — Текст» в направлениях синтеза или анализа сообщения.

## ЛИТЕРАТУРА

1. *Тр. Первого междунар. семинара «Компьютеризация естественных языков».* — Болгария, Варна, 3–7 сентября 1999 г. — Информационное обслуживание — АД. — 1999. — 147 с.
2. *Леонтьева Н.Н.* О предмете «прикладная лингвистика» (отвечая Н.В. Перцову) // Московский лингвистический альманах. — 1996. — Вып. 1. С. 234–244.
3. *Перцов Н.В.* О некоторых проблемах современной семантики и компьютерной лингвистики // Московский лингвистический альманах. — 1996. — Вып. 1. — С. 9–66.
4. *Кибрик А.Е.* О «невыполненных обещаниях» лингвистики 50 – 60 годов // Московский лингвистический альманах. — 1996. — Вып. 1. — С. 230–233.
5. *Щерба Л.В.* Языковая система и речевая деятельность. — Л.: Наука, 1974. — С. 230–233.
6. *Мельников Г.П.* Системология и языковые проблемы кибернетики. — М.: Энергия, 1979. — 368 с.

7. Астахова Л.И. Предложение и его членение (прагматика, семантика, синтаксис). — ДГУ, 1992. — 160 с.
8. Звегинцев В.А. Предложение и его отношение к языку и речи. — М.: МГУ, 1976. — 308 с.
9. Кубрякова Е.С., Шахнарович А.М., Сахарный Л.В. Человеческий фактор в языке: язык и порождение речи. — М.: Наука, 1991. — 239 с.
10. Апресян Ю.Д. Лексическая семантика. — М.: Наука, 1974. — 367 с.
11. Анисимов А.В., Марченко А.А. Система обработки текстов на естественном языке // Искусственный интеллект. — 2002. — № 4. — С. 157–163.
12. Анисимов А.В. Информатика, Творчество, Рекурсия. — Киев: Наук. думка, 1988. — 224 с.
13. Гладун В.П. Процессы формирования новых знаний. — София: СД «Педагог-6», 1994. — 189 с.
14. Зенкин Г.М., Петров А.П. Функциональная организация зрительного процесса и принцип гештальта // Интеллектуальные процессы и их моделирование. — М.: Наука, 1987. — 397 с.
15. Глезер В. Д. «Зрение и мышление». — Санкт-Петербург : Наука, 1993. — 284 с.
16. Кисленко Ю.І. Архітектура мови (лінгвістичне забезпечення інтелектуальних інтегрованих систем) : Учбовий посібник. — Київ: Віпол, 1998. — 343 с.
17. Кисленко Ю.И. Рекурсивный синтаксический анализатор // Наук. вісник кафедри ЮНЕСКО Київського держ. лінгв. ун-ту. — 2000. — Вип. 1. — Київ. — С. 157–164.
18. Кисленко Ю.І. Кількісні оцінки актуалізації базової структури // Искусственный интеллект. — 1999. — № 1. — С. 55–60.

Поступила 1.09.2003