

ОСНОВНЫЕ ПРИНЦИПЫ СИНТЕЗА ЛОГИКО-ЛИНГВИСТИЧЕСКИХ МОДЕЛЕЙ

Аннотация. Проанализированы недостатки существующих теорий семантического и лингвистического анализов текстов. Предложена формальная модель, учитывающая семантико-синтаксическую структуру предложений естественного языка. Приведены примеры построения логико-лингвистических моделей различных типов предложений, на основе которых сформулированы основные принципы синтеза логико-лингвистических моделей предложений естественного языка.

Ключевые слова: логико-лингвистическая модель, текстовая информация, естественный язык, семантическая структура, компьютерная лингвистика.

ВВЕДЕНИЕ

Одной из основных задач, которые ставятся перед учеными в сфере искусственного интеллекта, — научиться анализировать и воссоздавать с помощью вычислительных машин смысл действий, образов, текстов и т.д. В частности, в области аналитической обработки текстовой информации за последние годы создано большое количество программных продуктов, которые в той или иной степени охватывают область семантического анализа электронных документов [1].

Также осуществлено большое количество теоретических разработок в сфере компьютерной и структурной лингвистики. Так, И.Р. Гальперин в работе [2] дает полную характеристику внутри текстовых связей, раскрывая на примерах понятия континуума и членимости текста. Монография служит хорошим теоретическим пособием по лингвистическому анализу текста, но не имеет связи с автоматизированной обработкой текстовой информации. Еще одним источником знаний о тексте, без которых невозможна его компьютерная обработка, является курс лекций К.А. Филиппова [3], где детально представляется информация об организации текста, видах текста, а также способах связи между сложными частями документов. В монографии [4] Дж. Лайонз приводит всесторонний анализ текстов, обращая особое внимание на смысловое значение высказываний, предложений, приближаясь к понятию формализации знаний.

В.А. Широков в своей теории семантических состояний [5] определяет соответствие между языковой единицей и ее концептуальным представлением и предлагает общую форму представления семантического состояния единиц лексического уровня, используя фактор взаимодействия лексической и грамматической семантик. Работа коллектива авторов [6] раскрывает вопросы, начиная от теории лингвистического и математического моделирования и заканчивая вариантами технологических решений. В статье [7] разработана унифицированная форма представления знаний предложений естественного языка.

На сегодняшний день различают два типа алгоритмов определения вхождений в текстовые документы одних и тех же сущностей: алгоритмы, использующие тезаурус естественного языка, и разделенные алгоритмы, определяющие слова с одинаковым разделенным контекстом [8]. Также многие программные модули используют для анализа текста алгоритмы поиска в строке [9, 10].

Наиболее приближенным программным продуктом, который определяет семантические значения слов (понятия) в тексте и выделяет все связи между ними, создавая универсальное и не зависящее от языка представление текста [11], явля-

ется технология ABBYY Compeno. Анализ и поиск данных с помощью этой технологии проводится только на английском и русском языках.

Несмотря на огромную информационную базу (как в области лингвистики, так и в области информационных технологий) единого решения по смысловой автоматической обработке электронных документов до сих пор не выработано. В связи с этим остается актуальным вопрос о разработке моделей представления знаний, а также формализации процесса извлечения знаний из текстов разнообразной структуры.

ПОСТРОЕНИЕ ЛОГИКО-ЛИНГВИСТИЧЕСКИХ МОДЕЛЕЙ РАЗЛИЧНЫХ ТИПОВ ПРЕДЛОЖЕНИЙ

Для дальнейшего исследования связей между предложениями текста важно понять концепцию построения логико-лингвистических моделей тождественных по смыслу предложений. Также необходимо исследование различных типов предложений, от которых зависит изменение в логико-лингвистических моделях [12].

Каждому предложению присущи такие компоненты, как предикат (сказуемое), грамматический субъект (подлежащее) и логический субъект, которые задают грамматико-смысловую перспективу предложения. Исходя из этого, произвольное предложение естественного языка можно представить совокупностью простых предикатов, каждый из которых выражает наименьшую составляющую, имеющую законченный и самостоятельный смысл:

$$L_g(S) = \bigwedge_{\mu=1}^{v(S)} L_\mu(S), \quad (1)$$

где $L_\mu(S)$ — простой предикат, который описывает часть предложения S и отображает законченный смысл; $\mu = 1, v(S)$, $v(S)$ — количество частей, из которых состоит предложение S , отображающие законченный смысл.

Простой предикат записывается в виде формулы

$$L_\mu(S) = p(x_1, c(x_1), x_3, x_2, c(x_2), z, c(z), v(p)), \quad (2)$$

где x_1 — субъект предложения S , $x_1 \in X_1$; X_1 — множество субъектов, входящих в предложение S , $X_1 \subseteq X(S)$; $X(S)$ — множество сущностей, которые входят в предложение S ; $c(x_1)$ — кортеж характеристик субъекта x_1 :

$$c(x_1) = [c_k(x_1) | k = \overline{1, m_1(x_1)}];$$

$m_1(x_1)$ — количество характеристик субъекта x_1 ; x_2 — объект предложения S , $x_2 \in X_2(x_1)$; $X_2(x_1)$ — множество объектов, над которыми выполняет действие субъект x_1 ; $X_2(x_1) \subseteq X(S)$; $c(x_2)$ — кортеж характеристик объекта x_2 :

$$c(x_2) = [c_l(x_2) | l = \overline{1, m_2(x_2)}];$$

$m_2(x_2)$ — количество характеристик объекта x_2 ; x_3 — объект, который характеризует субъект x_1 предложения S , $x_3 \in X_3(x_1)$; $X_3(x_1)$ — множество объектов, которые характеризуют субъект x_1 , $X_3(x_1) \subseteq X(S)$, $x_3 \neq x_2$; p — отношение между субъектом x_1 и объектом x_2 предложения S , $p \in P(x_1, x_2)$; $P(x_1, x_2)$ — множество отношений между субъектом x_1 и объектом x_2 ; z — объект p -го отношения между субъектом x_1 и объектом x_2 , $z \in Z(x_1, x_2, p)$; $Z(x_1, x_2, p)$ — множество объектов p -го отношения между субъектом x_1 и объектом x_2 ; $c(z)$ — кортеж параметров z -го объекта p -го отношения между

субъектом x_1 и объектом x_2 :

$$c(z) = [c_q(z) \mid q = \overline{1, m_3(z)}];$$

$m_3(z)$ — количество параметров z -го объекта; $v(p)$ — кортеж параметров p -го отношения между субъектом x_1 и объектом x_2 :

$$v(p) = [v_i(p) \mid i = \overline{1, m(p)}];$$

$m(p)$ — количество параметров p -го отношения.

Таким образом, содержание предложения естественного языка будет развернуто в линейную, синтаксически правильную структуру. Рассмотрим возможные варианты построения тождественных по смыслу предложений естественного языка различной структуры.

Простое предложение без однородных членов. Пусть имеем следующее предложение: «*Отец оставил сыну в наследство дом*».

Логико-лингвистическая модель предложения:

$$L_1(S_1) = p'(x'_1, 0, 0, x'_2, 0, z', 0, w'(z')),$$

$$L_1(S_1) = \text{Оставил (отец, 0, 0, сыну, 0, наследство, 0, дом)}.$$

Предложение, тождественное по смыслу, можно представить с помощью перифраза в таком виде: «*Сын получил в наследство от отца дом*».

Логико-лингвистическая модель предложения:

$$L_2(S_2) = p''(x''_1, 0, 0, x''_2, 0, z'', 0, w''(z'')),$$

$$L_2(S_2) = \text{Получил (сын, 0, 0, отца, 0, наследство, 0, дом)}.$$

Еще один вариант записи тождественного по смыслу предложения: «*Дом был получен сыном в наследство от отца*». Такой вариант имеет место, поскольку при страдательном и действительном залогах грамматическое выражение одних и тех же отношений между производителем действия и объектом его действия осуществляется в параллельно противоположном направлении. Именно противоположным параллельным построением достигается смысловой эффект, пассивная конструкция производит не изменение в фактическом содержании, а лишь переоценку роли ее отдельных частей. Такая переоценка осуществляется чисто грамматическими средствами [13].

Логико-лингвистическая модель предложения:

$$L_3(S_3) = p'''(x'''_1, 0, 0, x'''_2, 0, z''', 0, w'''(z''')),$$

$$L_3(S_3) = \text{Был_получен (дом, 0, 0, отца, 0, сыном, 0, наследство)}.$$

Так как слова «*получил*» и «*был_получен*» являются активной и пассивной формами одного и того же глагола; существительные «отец» и «сын» принадлежат к классу гипонимов; глаголы «оставил» и «получил» являются конверсивами, т.е. имеют обратный смысл, поэтому можно сделать замены:

$$(p'' \equiv p') \& (x''_1 \equiv x'_2) \& (x''_2 \equiv x'_1), \quad z'' = z', \quad w''(z'') = w'(z'),$$

$$(p''' \equiv p'') \& (w'''(z''') \equiv z'') \& (x'''_1 \equiv w''(z'')) \& (x'''_2 \equiv x''_2) \& (x'''_1 \equiv x''_1).$$

В результате замены логико-лингвистические модели предложений будут иметь следующий вид:

$L_1(S_1) = p'(x'_1, 0, 0, x'_2, 0, z', 0, w'(z'))$, Оставил (отец, 0, 0, сыну, 0, наследство, 0, дом),

$L_2(S_2) = p'(x'_1, 0, 0, x'_2, 0, z', 0, w'(z'))$, Оставил (отец, 0, 0, сыну, 0, наследство, 0, дом),

$L_3(S_3) = p'(x'_1, 0, 0, x'_2, 0, z', 0, w'(z'))$, Оставил (отец, 0, 0, сыну, 0, наследство, 0, дом).

Таким образом, получены одинаковые логико-лингвистические модели для различных по грамматической структуре, но тождественных по смыслу предложений.

Сложносочиненное предложение с однородными членами. Пусть задано предложение «После проведения указанных оценок наступает ответственный момент окончательного и бесповоротного выбора единственной стратегии, и этот выбор осуществляется административными органами».

Логико-лингвистическая модель предложения сложносочиненного предложения:

$$\begin{aligned} L(S) &= (L'_1(S) \& L'_2(S) \& L'_3(S)) \& L''_2(S) = \\ &= (p'_1(x'_1, c'(x'_1), x'_3, x'_2, 0, z'_1, v'(p'_1), w'(z'_1)) \& \\ & p'_1(x'_1, c'(x'_1), 0, x'_3, c'_1(x'_3), z'_2, v'(p'_1), 0) \& \\ & p'_1(x'_1, c'(x'_1), 0, x'_3, c'_2(x'_3), z'_2, v'(p'_1), 0) \& \\ & p''_2(x'_3, c'(x'_3), 0, x'_2, c''(x'_2), 0, 0, 0), \end{aligned}$$

$$\begin{aligned} L(S) &= (L'_1(S) \& L'_2(S) \& L'_3(S)) \& L''_2(S) = \\ &= (\text{Наступает (момент, ответственный, выбора, проведения, 0, оценок,} \\ & \quad \text{указанных, после) \&} \\ & \quad \text{Наступает (момент, ответственный, 0, выбора, окончательного,} \\ & \quad \text{стратегии, единственной, 0) \&} \\ & \quad \text{Наступает (момент, ответственный, 0, выбора, бесповоротного,} \\ & \quad \text{стратегии, единственной, 0) \&} \\ & \quad \text{Осуществляется (выбор, этот, 0, органами, административными, 0, 0, 0).} \end{aligned}$$

Так как предложение состоит из двух частей, то комбинаций для формирования тождественных ему по смыслу предложений будет, как минимум,

$$C_3^2 = \frac{3!}{(3-2)!} = 6, \text{ если каждое из простых предложений интерпретировать тремя}$$

различными грамматическими параметрами.

Для каждого тождественного по смыслу предложения строятся логико-лингвистические модели, осуществляются замены согласно грамматическим параметрам, вследствие чего логико-лингвистические модели предложений приобретают одинаковый вид.

Сложноподчиненное предложение с деепричастным оборотом. Пусть имеется предложение «Достигнув уровня эксперта, человек делится опытом с учениками, чтобы передать им свои знания».

Логико-лингвистическая модель сложноподчиненного предложения с деепричастным оборотом:

$$\begin{aligned} L(S) &= L'_1(S) \rightarrow L''_2(S) \rightarrow L'''_3(S) = p'(x'_1, 0, 0, x'_2, 0, z', 0, 0) \rightarrow \\ & p''(x'_1, 0, 0, x'_2, 0, z'', 0, 0) \rightarrow p'''(0, 0, 0, z''', 0, z''', w'''(p'''), 0), \\ L(S) &= L'_1(S) \rightarrow L''_2(S) \rightarrow L'''_3(S) = \\ & = \text{Достиг (человек, 0, 0, уровня, 0, эксперта, 0, 0) \rightarrow} \end{aligned}$$

*Делится (человек, 0, 0, опытом, 0, учениками, 0, 0) →
 Передать (0, 0, 0, ученикам, 0, знания, свои, 0).*

Из примеров видно, что структура предложения, в частности наличие однородных членов, причастных и деепричастных оборотов, а также вид предложения, непосредственно влияет на количество компонентов логико-лингвистической модели и на операторы, которые в ней используются.

Рассмотрение различных по синтаксической структуре предложений, а также особенностей их формирования дает возможность излагать основные принципы синтеза логико-лингвистических моделей с целью формализовать процесс извлечения смысла из фрагментов текста.

ПРИНЦИПЫ СИНТЕЗА ЛОГИКО-ЛИНГВИСТИЧЕСКИХ МОДЕЛЕЙ

Синтез логико-лингвистических моделей — это объединение и замена их структурных компонентов благодаря выявлению способов логической связи между предложениями естественного языка в целях прослеживания единой тематической прогрессии текстового фрагмента.

Синтез логико-лингвистических моделей предполагает получение некоторой канонической формы, которая содержит элементарные отношения, имеющие тождественный смысл, и позволяет однозначно интерпретировать предложения естественного языка произвольной структуры.

Между структурным и семантическим планами нет тождества, но есть параллелизм, который проявляется в связях. Поскольку любая синтаксическая структура существует для передачи некоторого смысла, то назначение структурного плана — сделать возможным выражение мыслей, т.е. объектов семантического плана [13]. Поэтому произвольный текст можно представить в виде совокупности логико-лингвистических моделей предложений:

$$t' = \bigwedge_{g=1}^{N(t)} L_g(S_g), \quad (3)$$

где $L_g(S_g)$ — логико-лингвистическая модель предложения S_g , $g = \overline{1, N(t)}$; $N(t)$ — количество предложений в тексте t .

Логические связи в тексте можно определить как путем анализа сложного предложения (наличие деструктивных слов и анафорических связей в нем), так и выявлением лексико-грамматических характеристик (различным языковым отношениям между предложениями), которые обеспечивают формально-грамматическую связность текста, т.е. когезию [14].

Пусть имеется два предложения естественного языка: S_1 и S_2 , каждому из них соответствует логико-лингвистическая модель:

$$L_1(S_1) = p'_1(x'_1, c'(x'_1), x'_3, x'_2, c'(x'_2), z', v'(p'_1), w'(z')),$$

$$L_2(S_2) = p''_2(x''_1, c''(x''_1), x''_3, x''_2, c''(x''_2), z'', v''(p''_2), w''(z'')),$$

построенная согласно формулам (1), (2).

В целях получения формальной модели, с помощью которой можно описать фрагмент текста, учитывается смысл произвольных предложений естественного языка. Рассмотрим основные принципы синтеза логико-лингвистических моделей.

Принцип 1. Ему соответствует ситуация, когда для связи предложений в тексте используется семантическое повторение, при этом синтаксическая структура предложений одинаковая, но предикат, субъект или объект можно заменить синонимами, однокоренными словами или они могут повторяться.

1. Если хотя бы одно из отношений $p'_\mu \in P'_1(x'_1, x'_2)$ (где $P'_1(x'_1, x'_2)$ — множество отношений между субъектом x'_1 и объектом x'_2 предложения S_1) тождественно или является синонимом хотя бы одного из отношений $p''_\mu \in P''_2(x''_1, x''_2)$ предложения S_2 , то необходимо осуществить синтез, сделав в логико-лингвистических моделях $L_1(S_1)$ и $L_2(S_2)$ замену $p'_\mu = p''_\mu$:

$$L_1(S_1) = p'_1(x'_1, c'(x'_1), x'_3, x'_2, c'(x'_2), z', v'(p'_1), w'(z')),$$

$$L_2(S_2) = p'_1(x''_1, c''(x''_1), x''_3, x''_2, c''(x''_2), z'', v''(p''_1), w''(z'')).$$

После этого можно констатировать существование такого оператора $G_i(l_i)$, который ставит в соответствие вектору характеристик l_i каждого предложения S_i естественного языка определенную компоненту логико-лингвистической модели $u_k(S_j) \in U$ предложения S_j естественного языка, связанного по смыслу с предложением S_i :

$$G_i(l_i): U \rightarrow u_k(S_j),$$

где U — множество возможных значений компонент логико-лингвистической модели (2).

Согласно принципу 1 (п. 1) вектору l_1 характеристик предложения S_1 естественного языка ставится в соответствие значение $u_k(S_j) = u_1(S_2) = p'_1$, а вектору l_2 характеристик предложения S_2 естественного языка ставится в соответствие значение $u_k(S_j) = u_1(S_1) = p''_2$.

Рассмотрим фрагмент текста, состоящий из двух предложений: «*Англичане очень любят спортивные соревнования*» и «*Они любят заниматься спортом, смотреть соревнования или обсуждать их*».

Логико-лингвистическая модель фрагмента имеет следующий вид:

$$L_1(S_1) = p'_1(x'_1, 0, 0, x'_2, c'(x'_2), 0, 0, w'(z')),$$

$$L_2(S_2) = (p'_1 \& p''_{21})(x'_1, 0, 0, x''_{21}, 0, 0, 0, 0) \& ((p'_1 \& p''_{22})(x'_1, 0, 0, x''_{22}, 0, 0, 0, 0) \vee \\ \vee ((p'_1 \& p''_{23})(x'_1, 0, 0, x''_{22}, 0, 0, 0, 0)),$$

$$L_1(S_1) = \text{Любят (англичане, 0, 0, соревнования, спортивные, 0, 0, очень)},$$

$$L_2(S_2) = \text{Любят \& заниматься (англичане, 0, 0, спортом, 0, 0, 0, 0) \&}$$

$$\text{(Любят \& смотреть (англичане, 0, 0, соревнования, 0, 0, 0, 0) \vee}$$

$$\text{Любят \& обсуждать (англичане, 0, 0, соревнования, 0, 0, 0, 0))}.$$

2. Если субъект x'_1 предложения S_1 , $x'_1 \in X'_1$, где X'_1 — множество субъектов предложения S_1 , является синонимом, однокоренным словом или тождественно равен субъекту x''_1 либо объекту x''_2 предложения S_2 , $x''_1 \in X''_1$, $x''_2 \in X''(x''_1)$, где X''_1 — множество субъектов предложения S_2 , а X'' — множество сущностей предложения S_2 , то в результате синтеза будут получены логико-лингвистические модели, в которых $(x''_1 = x'_1) \vee (x''_2 = x'_1)$. Соответственно кортеж характеристик будет также соответствовать заменяемой сущности:

$$L_1(S_1) = p'_1(x'_1, c'(x'_1), x'_3, x'_2, c'(x'_2), z', v'(p'_1), w'(z')),$$

$$L_2(S_2) = p''_2(x'_1, c'(x'_1), x''_3, x''_2, c''(x''_2), z'', v''(p''_2), w''(z''))$$

или

$$L_1(S_1) = p'_1(x'_1, c'(x'_1), x'_3, x'_2, c'(x'_2), z', v'(p'_1), w'(z')),$$

$$L_2(S_2) = p''_2(x''_1, c''(x''_1), x''_3, x'_1, c'(x'_1), z'', v''(p''_2), w''(z'')).$$

Вектору l_1 характеристик предложения S_1 естественного языка ставится в соответствие значение $u_k(S_j) = u_2(S_2) = x_1'' \& c''(x_1'')$ или $u_k(S_j) = u_4(S_2) = x_2'' \& c''(x_2'')$, а вектору l_2 характеристик предложения S_2 естественного языка ставится в соответствие значение $u_k(S_j) = u_2(S_1) = x_1' \& c'(x_1')$.

Рассмотрим фрагмент текста, состоящий из двух предложений: «Испокон веков британской игрой считался крокет» и «Крокет олицетворяет дух Англии».

Логико-лингвистическая модель фрагмента имеет следующий вид:

$$L_1(S_1) = p_1'(x_1', 0, 0, x_2', c'(x_2'), 0, 0, w'(z')),$$

$$L_2(S_2) = p_2''(x_1'', 0, 0, x_2'', 0, z'', 0, 0),$$

$$L_1(S_1) = \text{Считался (крокет, 0, 0, игрой, британской, 0, 0, испокон веков),}$$

$$L_2(S_2) = \text{Олицетворяет (крокет, 0, 0, дух, 0, Англии, 0, 0).}$$

3. Синтез осуществляется аналогично п. 2, но по отношению к равенству субъекта x_2'' второго предложения. Тогда логико-лингвистические модели приобретают вид

$$L_1(S_1) = p_1'(x_2'', c''(x_2''), x_3', x_2', c'(x_2'), z', v'(p_1'), w'(z')),$$

$$L_2(S_2) = p_2''(x_1'', c''(x_1''), x_3'', x_2'', c''(x_2''), z'', v''(p_2''), w''(z''))$$

или

$$L_1(S_1) = p_1'(x_1', c'(x_1'), x_3', x_2'', c''(x_2''), z', v'(p_1'), w'(z')),$$

$$L_2(S_2) = p_2''(x_1'', c''(x_1''), x_3'', x_2'', c''(x_2''), z'', v''(p_2''), w''(z'')).$$

Вектору характеристик l_1 предложения естественного языка S_1 ставится в соответствие значение $u_k(S_j) = u_4(S_2) = x_2'' \& c''(x_2'')$, а вектору характеристик l_2 предложения естественного языка S_2 ставится в соответствие значение $u_k(S_j) = u_2(S_1) = x_1' \& c'(x_1')$ или $u_k(S_j) = u_4(S_1) = x_2' \& c'(x_2')$.

Рассмотрим фрагмент текста, состоящий из двух предложений: «Испокон веков британской игрой считался крокет» и «В крокет играют в школах, колледжах, университетах».

Логико-лингвистическая модель фрагмента имеет следующий вид:

$$L_1(S_1) = p_1'(x_2'', 0, 0, x_2', c'(x_2'), 0, 0, w'(z')),$$

$$L_1(S_1) = p_2''(0, 0, 0, x_2'', 0, z_1'', 0, 0) \& p_2''(0, 0, 0, x_2'', 0, z_2'', 0, 0) \&$$

$$p_2''(0, 0, 0, x_2'', 0, z_3'', 0, 0),$$

$$L_1(S_1) = \text{Считался (крокет, 0, 0, игрой, британской, 0, 0, испокон веков),}$$

$$L_2(S_2) = \text{Играют (0, 0, 0, крокет, 0, школах, 0, 0) \&}$$

$$\text{Играют (0, 0, 0, крокет, 0, колледжах, 0, 0) \&}$$

$$\text{Играют (0, 0, 0, крокет, 0, университетах, 0, 0).}$$

Принцип 2. Принцип синтеза логико-лингвистических моделей базируется на дейктическом повторении. Это один из способов контактной связи в текстах и означает, что на уже названные предметы, характеристики и обстоятельства ссылаются в следующих предложениях, используя такие специальные слова, как, например, местоимения, наречия.

1. Если субъект x_1'' или объект x_2'' предложения S_2 ($x_1'' \in X_1''$, $x_2'' \in X''(x_1'')$, где X_1'' — множество субъектов предложения S_2 , а X'' — множество сущностей предложения S_2), является местоимением, то вследствие синтеза они будут заменены на элемент множества X' (сущностей предложения S_1), грамматические характеристики которого совпадают с характеристиками x_1'' или x_2'' . При этом кор-

теж характеристик субъекта x_1'' или объекта x_2'' предложения S_2 будет сохраняться прежним. Таким образом, в результате синтеза двух предложений будут получены логико-лингвистические модели, в которых $(x_1'' = x_1') \vee (x_2'' = x_1')$:

$$L_1(S_1) = p_1'(x_1', c'(x_1'), x_3', x_2', c'(x_2'), z', v'(p_1'), w'(z')),$$

$$L_2(S_2) = p_2''(x_1'', c''(x_1''), x_3'', x_2'', c''(x_2''), z'', v''(p_2''), w''(z''))$$

или

$$L_1(S_1) = p_1'(x_1', c'(x_1'), x_3', x_2', c'(x_2'), z', v'(p_1'), w'(z')),$$

$$L_2(S_2) = p_2''(x_1'', c''(x_1''), x_3'', x_1', c''(x_2''), z'', v''(p_2''), w''(z'')).$$

Вектору характеристик l_1 предложения S_1 естественного языка ставится в соответствие значение $u_k(S_j) = u_2(S_2) = x_1''$ или $u_k(S_j) = u_4(S_2) = x_2''$, а вектору характеристик l_2 предложения S_2 естественного языка ставится в соответствие значение $u_k(S_j) = u_2(S_1) = x_1'$.

Рассмотрим фрагмент текста, состоящий из таких предложений: «*Англичане очень любят спортивные соревнования*» и «*Они занимаются спортом*».

Логико-лингвистическая модель фрагмента имеет следующий вид:

$$L_1(S_1) = p_1'(x_1', 0, 0, x_2', c'(x_2'), 0, 0, w'(z')),$$

$$L_2(S_2) = p_2''(x_1'', 0, 0, x_2'', 0, 0, 0, 0),$$

$$L_1(S_1) = \text{Любят (англичане, 0, 0, соревнования, спортивные, 0, 0, очень)},$$

$$L_2(S_2) = \text{Занимаются (англичане, 0, 0, спортом, 0, 0, 0, 0)}.$$

2. Если среди параметров кортежа отношений $v''(p_2'')$ выявлены наречия в виде слов-указателей, то составляющие $v'(p_1')$ и $v''(p_2'')$ можно отождествить, выполнив синтез логико-лингвистических моделей предложений, в которых $v''(p_2'') = v'(p_1')$:

$$L_1(S_1) = p_1'(x_1', c'(x_1'), x_3', x_2', c'(x_2'), z', v'(p_1'), w'(z')),$$

$$L_2(S_2) = p_2''(x_1'', c''(x_1''), x_3'', x_2'', c''(x_2''), z'', v''(p_2''), w''(z'')).$$

Вектору характеристик l_1 предложения S_1 естественного языка ставится в соответствие значение $u_k(S_j) = u_6(S_2) = v''(p_2'')$, а вектору характеристик l_2 предложения S_2 естественного языка ставится в соответствие значение $u_k(S_j) = u_6(S_1) = v'(p_1')$.

Рассмотрим фрагмент текста, состоящий из двух предложений: «*Эксперты озвучили продолжительность пребывания на Марсе*» и «*Там проводились эксперименты на животных*».

Логико-лингвистическая модель фрагмента имеет следующий вид:

$$L_1(S_1) = p_1'(x_1', 0, 0, x_2', 0, z', 0, w'(z')),$$

$$L_2(S_2) = p_2''(x_1'', 0, 0, x_2'', 0, 0, 0, w''(z'')),$$

$$L_1(S_1) = \text{Озвучили (эксперты, 0, 0, продолжительность, 0, пребывания, 0, Марсе)},$$

$$L_2(S_2) = \text{Проводились (эксперименты, 0, 0, животных, 0, 0, 0, Марсе)}.$$

Принцип 3. Принцип формулируется, исходя из повторения грамматических значений в предложениях. Так, в связном тексте глагольные формы объединяют предложения естественного языка общим временным планом, несмотря на то, что как предикаты, так и объекты двух предложений не являются тождественными. В этом случае синтез логико-лингвистических моделей будет иметь

следующий вид:

$$L_1(S_1) = p'_1(x'_1, c'(x'_1), x'_3, x'_2, c'(x'_2), z', v'(p'_1), w'(z')),$$

$$L_2(S_2) = p''_2(x''_1, c''(x''_1), x''_3, x''_2, c''(x''_2), z'', v''(p''_2), w''(z'')).$$

Вектору характеристик l_1 предложения S_1 естественного языка ставится в соответствие значение $u_k(S_j) = u_1(S_2) = p'_1$, а вектору характеристик l_2 предложения S_2 естественного языка ставится в соответствие значение $u_k(S_j) = u_1(S_1) = p''_2$.

Рассмотрим фрагмент текста, состоящий из двух предложений: «*Испокон веков британской игрой считался крокет*» и «*Футбол являлся самой популярной игрой в Англии*».

Логико-лингвистическая модель фрагмента имеет следующий вид:

$$L_1(S_1) = p'_1(x''_2, 0, 0, x'_2, c'(x'_2), 0, 0, w'(z')),$$

$$L_2(S_2) = p''_2(x''_1, 0, 0, x''_2, c''(x''_2), 0, 0, w''(z'')),$$

$$L_1(S_1) = \text{Считался (крокет, 0, 0, игрой, британской, 0, 0, испокон веков),}$$

$$L_2(S_2) = \text{Являлся (футбол, 0, 0, игрой, самой популярной, 0, 0, Англии).}$$

Принцип 4. Принцип базируется на замене синонимов как из одной, так и разных частей речи. Это становится возможным благодаря синтаксической или транс-позиционной деривации: переходе слова из одной части речи в другую без замены их лексических значений [15]. Наиболее распространенные способы построения такого рода синонимических конструкций: «глагол1» — «глагол2 + однокоренное существительное глагола1», «глагол1 — однокоренное существительное глагола1» и «глагол1 — однокоренное прилагательное глагола1». Приведением глагольной формы предикатов к эквивалентным существительным можно осуществить синтез логико-лингвистических моделей, при котором субъект x''_1 предложения S_2 заменяется предикатом $p''_1 \in P'_1(x'_1, x'_2)$ предложения S_1 , при этом объекты $x''_2 = x'_2$ предложений также равны согласно принципу 2 синтеза:

$$L_1(S_1) = p'_1(x'_1, c'(x'_1), x'_3, x'_2, c'(x'_2), z', v'(p'_1), w'(z')),$$

$$L_2(S_2) = p''_2(p'_1, 0, x''_3, x'_2, c''(x''_2), z'', v''(p''_2), w''(z'')).$$

Вектору характеристик l_1 предложения S_1 естественного языка ставится в соответствие значение $u_k(S_j) = u_2(S_2) = x''_1 \& c''(x''_1) \& x''_2$, а вектору характеристик l_2 предложения S_2 естественного языка ставится в соответствие значение $u_k(S_j) = u_1(S_1) = p'_1 \& x'_2$.

Рассмотрим фрагмент текста, состоящий из двух предложений: «*Книга заинтересовала любителей приключений*» и «*Книга интересна любителям приключений*».

Логико-лингвистическая модель фрагмента имеет следующий вид:

$$L_1(S_1) = p'_1(x'_1, 0, 0, x'_2, 0, z', 0, 0),$$

$$L_2(S_2) = p''_2(x'_1, 0, 0, x'_2, 0, z', 0, 0),$$

$$L_1(S_1) = \text{Заинтересовала (книга, 0, 0, любителей, 0, приключений, 0, 0),}$$

$$L_2(S_2) = \text{Интересна (книга, 0, 0, любителям, 0, приключений, 0, 0).}$$

Применяя принципы синтеза логико-лингвистических моделей рекурсивно и проходя все цепочки связей от каждого из предложений, можно получить конечное множество моделей t' , выражающих все элементы смысла.

Сформулированные выше принципы синтеза логико-лингвистических моделей используются для формирования множества продукций D текстовой базы.

Такое множество содержит модели $L'_\mu(S)$, которые по смыслу предшествуют модели $L''_\mu(S)$, т.е. $L'_\mu(S) \rightarrow L''_\mu(S)$. На основании таких связей строится общая содержательная модель текста.

ЗАКЛЮЧЕНИЕ

Основные компоненты представления смысла предложения — это лексический состав и синтаксическая структура. Синтаксис задает на множестве лексем отношение подчинения, по которому можно выявить предикатно-аргументные отношения между оператором и операндом в семантической структуре [16]. Предложенная в статье форма (1)–(2) построения логико-лингвистических моделей предложений естественного языка учитывает их семантико-синтаксическую структуру, что позволяет формально описывать смысл текстовой информации.

Синтез логико-лингвистических моделей предложений естественного языка является неотъемлемой составляющей процесса лингвистического анализа текста в целом. Только с помощью синтеза логико-лингвистических моделей можно построить формальную модель текста. Дальнейшие исследования направлены на создание метода автоматизированного построения логико-лингвистических моделей электронных текстов в целях их сравнения по смыслу и поиска логических противоречий.

СПИСОК ЛИТЕРАТУРЫ

1. Каталог лингвистических программ и ресурсов в Сети — <http://www.rvb.ru/soft/catalogue/catalogue.html>.
2. Гальперин И.Р. Текст как объект лингвистического исследования. Изд. 5-е, стереотипное. — М.: КомКнига, 2007. — 144 с.
3. Филиппов К.А. Лингвистика текста. Курс лекций. — СПб.: Изд-во С.-Пб ун-та, 2008. — 336 с.
4. Лайонз Дж. Лингвистическая семантика. — М.: Языки славянской культуры, 2003. — 400 с.
5. Лінгвістичні та технологічні основи тлумачної лексикографії / В.А. Широков, В.М. Білоноженко, О.В. Бугаков та ін.— Київ: Довіра, 2010. — 295 с.
6. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика / Е.И. Большакова, Э.С. Клышинский, Д.В. Ландэ, А.А. Носков, О.В. Пескова, Е.В. Ягунова. — М.: МИЭМ, 2011. — 272 с.
7. Вавіленкова А.І. Логіко-лінгвістична модель як засіб відображення синтаксичних особливостей текстової інформації // Математичні машини та системи. — 2010. — № 2. — С. 134–137.
8. Jurafsky D., Manning Ch. Natural language processing, 2012. — <https://www.coursera.org/course/nlp> (accessed 15.05.2014).
9. Поддержка синонимов в программах поиска. — http://www-01.ibm.com/support/knowledgecenter/SSEPGG_8.2.0/com.ibm.db2.ii.of.doc/admin/iisysasynsupp.htm?lang=ru.
10. Русский Wordnet. — <http://wordnet.ru/>.
11. АBBYU Intelligent Search SDK. — <http://www.abbyu.ru/isearch/compreno/>.
12. Вавіленкова А.І. Методика перетворення речення в логіко-лінгвістичну модель // Міжнарод. наук.-техн. конф. «Комп'ютерні системи та мережні технології». — Київ: НАУ, 2008. — С. 49–53.
13. Теньер Л. Основы структурного синтаксиса: пер.с франц. — М.: Прогресс, 1988. — 656 с.
14. Лукашевич Н.В. Когезия как структурная связность текста, 2011. — <http://like-money.ru/stati/271-kogeziya-kak-strukturnaya-svyaznost-teksta>.
15. Кормен Т. Алгоритмы: построение и анализ. Изд. 3-е / Т. Кормен, И. Чарльз, Р. Лейзерсон, Л. Ривест, К. Штайн. — СПб.: Вильямс, 2013. — 1328 с.
16. Падучева Е.В. Семантические исследования. — М.: Языки русской культуры, 1996. — 464 с.

Поступила 06.08.2014