

ФОРМАЛИЗАЦИЯ СТРУКТУРНО-СЕМАНТИЧЕСКОЙ ОРГАНИЗАЦИИ ЗАГОЛОВКОВ АНГЛОЯЗЫЧНЫХ НАУЧНЫХ СТАТЕЙ

Формализация семантико-синтаксической организации фрагмента текста предполагает использование некоторого искусственного языка, который мог бы понять компьютер. Проблема создания такого языка теснейшим образом связана с задачей моделирования естественного языка (ЕЯ). Выделяют три подхода к такому моделированию [1, с. 89–91]. При **первом подходе** устанавливаются однозначные соотношения между синтаксической структурой выражения на естественном языке и элементами смысла. Определенным лексическим единицам «приписываются» конкретные структуры синтаксических конструкций, которые, в свою очередь, и отражают смысл. При таком подходе сохраняется разрыв между языком и внешним предметным миром, поскольку указанная связь между лексическими единицами и элементами смысла устанавливается без привлечения лексических значений единиц ЕЯ. Этот принцип нашел отражение в алгоритмических языках, использующих лексику естественных языков. **Второй подход** к использованию ЕЯ для обработки текстов компьютером связан с анализом понятийной сферы внешнего предметного мира, расчленяемого на кванты смысла, которые отражаются конечной фиксированной совокупностью обобщенных лексических единиц. С помощью последних в правильные предложения разворачиваются единицы лексико-грамматических категорий. К числу таких языков можно отнести язык КХ кодов, язык «смысл–текст», язык УСК. Сюда можно отнести семейство языков для представления знаний. Общим для этих языков является то, что они являются средством представления значений отдельного слова, простого предложения или некоторого факта. Дальнейшее развитие этой группы языков сводится к тому, чтобы представить в удобном для компьютера виде сложные взаимосвязанные факты действительности, описанные связными текстами. **Третий подход** к моделированию ЕЯ с целью использования его для общения с компьютером сводится к статистическому выбору из текстов естественного языка, относящихся к узкой предметной области, наиболее употребительных лексических единиц и грамматических правил. При этом создается так называемый базовый язык, являющийся средством передачи знаний человеку или компьютеру о сугубо узкой предметной области. Сюда можно отнести такие языки как МЕДИФОР, PLANNER, а также ряд других. Как показывает опыт эксплуатации этих и подобных им языков, проблема общения человека с компьютером на естественном языке еще далека от полного решения.

Один из возможных подходов к созданию искусственного языка для представления знаний в памяти компьютера описан в работе [1, с. 93–139]. Автор разработал специальный язык СЕМСИНТ, который дает возможность описать содержание отдельного словосочетания, предложения, абзаца и текста в

целом. СЕМСИНТ позволяет строить семантико-синтаксическую формулу абзаца, а также содержит правила, позволяющие учитывать семантические и синтаксические отношения между текстовыми единицами, т. е. строить их семантико-синтаксические формулы. Данный язык включает следующие составляющие: 1) алфавит; 2) набор элементов для записи семантических отношений между членами словосочетания или предложения; 3) набор элементов для записи синтаксических отношений между этими текстовыми единицами.

Для описания структурно-семантической организации отобранных заголовков англоязычных научных текстов были использованы следующие семантические функции языка СЕМСИНТ:

Таблица 1

Использованные в работе семантические функции языка СЕМСИНТ

№ п/п	Код семантической функции	Наименование семантической функции	Содержание семантической функции
1.	AS2	Понятие	Характеризует аргументную группу как отражение в словосочетании <i>активного неодушевленного инициатора события (общего понятия)</i> вместе с его признаками.
2.	AS3	Процесс	Характеризует аргументную группу как отражение в словосочетании <i>активного неодушевленного инициатора события (процесса)</i> вместе с его признаками.
3.	AEL	П-деятель	Характеризует аргументную группу как отражение в словосочетании <i>активного неодушевленного инициатора события (природного или стихийного явления или предмета)</i> вместе с его признаками.
4.	AP	Объект	Характеризует аргументную группу как отражение в словосочетании <i>неактивного одушевленного объекта события</i> вместе с его признаками.
5.	AB1	Получатель	Характеризует аргументную группу как отражение в словосочетании <i>неактивного одушевленного адресата (получателя некоторого события)</i> .
6.	AB2	Адресат	Характеризует аргументную группу как отражение в словосочетании <i>неактивного неодушевленного адресата некоторого события</i> .
7.	AO	Н-объект	Характеризует аргументную группу как отражение в словосочетании <i>неактивного неодушевленного объекта некоторого события</i> .
8.	ALK	Место	Характеризует аргументную группу как отражение в словосочетании <i>места совершения некоторого события</i> .

№ п/п	Код семантической функции	Наименование семантической функции	Содержание семантической функции
9.	AMD	Средство	Характеризует аргументную группу как отражение в словосочетании <i>средства выполнения события</i> .
10.	AIN	Инструмент	Характеризует аргументную группу как отражение в словосочетании <i>орудия или инструмента выполнения события</i> .
11.	AKM	Состав	Характеризует аргументную группу как отражение в словосочетании <i>состава или содержимого какого-либо предмета</i> , выраженного другой аргументной группой.
12.	AD1	Определитель	Характеризует аргументную группу как отражение в словосочетании некоторого <i>свойства какого-либо предмета</i> , выраженного другой аргументной группой.
13.	ATM	Время	Характеризует аргументную группу как отражение в словосочетании <i>времени осуществления события</i> .
14.	AAD	Способ	Характеризует аргументную группу как отражение в словосочетании <i>способа реализации события</i> .
15.	AIF	Условие	Характеризует аргументную группу как отражение в словосочетании <i>условия осуществления события</i> .
16.	ACS	Причина	Характеризует аргументную группу как отражение в словосочетании <i>причины осуществления события</i> .
17.	AAM	Цель	Характеризует аргументную группу как отражение в словосочетании <i>цели выполнения события</i> .

Для увеличения точности отражения семантико-синтаксической формулой словосочетания элементов ситуации в язык СЕМСИНТ дополнительно введены семантические подклассы знаменательных классов слов. Условные обозначения этих подклассов приведены в табл.2.

Таблица 2

Коды семантических подклассов слов, использованных в языке СЕМСИНТ

№ п/п	Наименование семантического подкласса	Код семантического подкласса
1.	Существительное	NIJ
2.	Прилагательное	JIJ
3.	Местоимение	PIJ
4.	Наречие	DIJ
5.	Причастие	SIJ
6.	Числительное	MIJ
7.	Предлог	KIJ

Таким образом, аргументная группа в семантико-синтаксической формуле заголовка состоит из цепочки кодов семантических подклассов определителей и существительных, соединенных знаком * («звездочка»), ограниченных слева

знаком <, а справа знаком >. Эти знаки ставятся за кодами соответствующих подклассов частей речи. Например, заголовок *Speech Recognition System for the Russian Speech* будет представлен на языке СЕМСИНТ следующей семантико-синтаксической формулой:

AS2<N03*N04*N19>+FOR< K01>+AS2<J04*N03>,

где AS2 – код семантической функции «Понятие»;

N03 – семантический подкласс существительных, связанных с языком, языковым явлением;

N04 – семантический подкласс существительных, связанных с процессом выполнения действия при исследовании языкового явления;

N19 – семантический подкласс существительных, связанных со структурной организацией;

FOR – ядерный предлог словосочетания;

K01 – семантический подкласс неядерных предлогов;

J04 – семантический подкласс прилагательных, связанных с названиями языков.

Семантико-синтаксические формулы исследованных заголовков составили лингвистическую базу знаний системы автоматического порождения заголовка к научному тексту.

Литература

1. Зубов, А.В. Вероятностно-алгоритмическая модель порождения текста (семантико-синтаксический аспект): автореф. дис. ... д-ра фил. наук: 10.02.21. – М., 1985.