



Тартуский Государственный
Университет

ТЕЗИСЫ ДОКЛАДОВ СЕМИНАР ПРОЕКТА ДИАЛОГ

САНГАСТЕ 1-6 марта 1982 г.

ТАРТУ 1982

ТАРТУСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

СЕМИНАР ПРОЕКТА «ДИАЛОГ»

Тезисы докладов

Сангасте, 1-6 марта 1982 г.

ТАРТУ 1982

Оргкомитет семинара

А.П.Ершов - председатель
Д.А.Бухштаб - секретарь
В.М.Андрющенко
В.А.Брехов
О.С.Кулагина
Л.И.Микулич
А.С.Нариньяни
Д.А.Поспелов
М.Э.Салувээр
И.С.Сильдмэ

RASTUTATUD

Arh.
Tartu Riikliku Ülikooli
Raamatukogu

4428

СИСТЕМА МОРФОЛОГИЧЕСКОЙ ИНТЕРПРЕТАЦИИ ТЕКСТА

В.З. Демьянков

Под морфологической интерпретацией (МИ) понимается соотнесение текстовой и словарной форм лексемы при установлении грамматических помет словоформы. Система МИ позволяет: а) пользоваться в рамках человека-машинного общения стандартными словарями, б) определять морфный состав и семантику неологизмов (т.е. единиц, не содержащихся в словаре реальных единиц, но понятных носителю языка), в) конструировать толкования как хранимых, так и новых единиц словаря, г) автоматически кодировать текстовую информацию поморфемно, а не побуквенно. Она может быть использована также при синтезе текстовых форм и при распознавании устной речи. Система основана на принципах гипотетической интерпретации, лежащих в основе естественного интеллекта, в частности, на следующих:

1. МИ состоит в выяснении гипотетического морфемного состава основы у анализируемой словоформы (идя от более ранних по времени появления в тексте единиц к более поздним) и ее морфологического класса (того набора флексий со своими грамматическими пометами, которые к этой основе могут присоединяться); после чего гипотеза верифицируется (т.е. устанавливается, идет ли после основы флексия из предсказываемого множества).
2. Словарь содержит информацию о продуктивном материале для основ, а не сами основы; реальные и потенциальные основы языка конструируются из морфов, входящих в инвентарь системы МИ.
3. Чередования основ отражены в словаре как чередования соответствующих морфов.
4. Лексическое значение словоформы определяется как результат семантической интерпретации, основанной на окончательной или

промежуточной МИ. Семантические правила оперируют морфемами как элементарными единицами и делятся на два класса: продуктивные и уникальные. Первые устанавливают буквальное значение лексем (ее "внутреннюю форму"), вторые - все реальные значения.

К общим принципам гипотетической интерпретации относятся следующие: 1) процесс интерпретации - это постепенное расширение и сужение текущего набора гипотез, 2) интерпретация целого выражения базируется на интерпретации составных частей, а этапность интерпретации определяется конструкцией выражения, 3) интерпретация неправильно построенных выражений содержит, кроме прочего, указания на отклонения, релевантные для конкретного вида интерпретации. Система МИ основана на формальном аппарате рекурсивных функций при древовидном упорядочении инвентаря морфем. Этот инвентарь - словарь морфов, или лексикон - состоит из отдельных "статей", в каждой из которых при заглавном морфе указаны: 1) схема преобразования, по которому выражение, находящееся в поле зрения соответствующей функции (напр., функции идентификации корня, суффикса, флексии и т.п.), преобразуется в рамках текущей гипотезы, 2) конкретные "наполнители" схемы, специфичные для конкретного морфа. К наполнителям относится, напр., альтернант, т.е. то выражение, которое, заменив данный морф, приводит к имени лексемы. Альтернант флексии - это флексия "заглавного слова" в рамках сложившейся национальной лексикографической традиции с указанием грамматических признаков, вносимых в текстовую форму заменяемой флексией. Другой вид наполнителей - пометы, наличие которых является условием для работы преобразований и которые вводятся в выражение в результате предшествующих преобразований. Словарные статьи упорядочены так, что по объективному выражению, находящемуся в поле зрения ведущей функции, без лишних просмотров словаря выявляются все те морфы, которые могут начинать это выражение. Если ни одного такого морфа не найдено, значит гипотеза с таким "остатком" является посторонней и стирается. Если же таких морфов найдено несколько, то текущая гипотеза "расщепляется" на несколько новых, каждая из которых проверяется, начиная с самой левой в линейном представлении общего набора текущих гипотез. Поскольку длина морфа в среднем не превышает 3-4 фонем, расщепление гипотез не приводит к нереалистичному увеличению длины такого представления. С дру-

гой стороны, внутри отдельного "куста" морфем в словаре единицы и схемы преобразований упорядочены так, что более левые гипотезы — наиболее вероятные; поэтому сначала получаются на выходе наиболее "легкие" варианты интерпретации, а затем, пока частичный результат выпечатывается, — и остальные варианты.

В докладе показывается, как с помощью такого формального аппарата отражаются следующие типовые явления морфологии естественных языков: а) чередования (как внутренние сандхи, типа рус. "бегу" — "бежишь", так и внешние, кельтского типа), б) "разрозненность" в тексте частей словоформы (случаи типа отделяемых префиксов в нем., венг. и др.), в) продуктивность "сложения основ" (типа нем., финского, венг., латышского и др.), г) местоименная энклитизация (типа исп., итал. и др.), д) разрыв флексий (как в арабском), е) сингармонизм и агглютинация (туркского и финно-угорского типов), ж) удвоение (как типа англ., где оно является орфографическим приемом, так и типа венг., где оно представляет собой морфологическое словоизменяющее средство) и другие. Попутным продуктом МИ является установление способа прочтения словоформы, особенно важное для тех языков, в которых прочтение не прямо вытекает из графики, а определяется морфологическими свойствами (установление огласовок в семитских текстах, вариантов гласного "е" — закрытого или открытого — в латышском, акцентуация в языках типа рус., литовского и др., а также фонематическая транскрипция в японском иероглифическом тексте).

Принципы МИ и предлагаемый формальный аппарат приложимы также и на синтаксическом, семантическом, прагматическом и др. уровнях анализа текста: на каждом из них единообразно обращение к инвентарю хранимых языковых единиц, расширение и сужение множества гипотетических интерпретаций и сканирование текста.

СОДЕРЖАНИЕ

Ю.Д.Апресян. Лингвистические проблемы семантического анализа предложений естественного языка.	3
Ю.Д.Апресян, Е.В.Гецелевич, Л.П.Крысин, А.В.Лазурский, Н.В.Перцов, В.З.Санников. Толково-комбинаторный словарь как компонент семантического анализа текста.	7
О.Ю.Богуславская, И.А.Муравьева. Об одном аспекте изучения анафории.	II
И.М.Богуславский, Л.Л.Иомдин, Л.Л.Цинман. Опыт экспериментальной эксплуатации системы автоматического перевода ЭТАП-1 и разработка системы ЭТАП-2.	13
И.А.Большаков. Программа перевода с русского флективного Паскаля на английский.	16
В.Э.Вольфенгаген, А.В.Ганночка, М.Ю.Чуприков. Принципы представления знаний о языке диалога.	19
Б.С.Галышков. Разрывные синтаксические группы и полуконтекстные грамматики.	22
В.З.Демьянков. Система морфологической интерпретации текста.	23
А.Я.Диковский, А.В.Гладкий. Об одном подходе к созданию точного языка семантики.	26
Н.В.Ершова. Построение связанных текстов в диалоговой системе с фреймовым представлением знаний.	28
Е.Р.Иоанесян, Е.С.Кузьмина, А.В.Лазурский, В.Ю.Розенцвейг, О.А.Штернова. Словарь глагольных валентностей на материале французского языка.	31
О.А.Казакевич. К вопросу о причинном анализе текста.	34
А.Е.Кибрик. Компоненты семантического представления.	37
М.Э.Койт. Морфологический синтезатор системы TARTUS.	40
И.С.Конonenко, Е.Л.Першина. Некоторые глубинно-синтаксические отношения для системы синтеза русского текста.	42
Б.М.Лейкина. О нескольких фрагментах базовой грамматики содержания.	45
С.Р.Литвак. Система ПРИЗ как основа построения системы представления знаний.	47

В.С.Лозовский. Неопределённость в системах представления знаний о мире.	50
М.Г.Мальковский, М.Ю.Семёнов. Использование векторов для синтаксического анализа фраз естественного языка.	53
М.Б.Мацкин. Интеллектуальные средства отладки задач.	56
А.С.Нариньяни, Е.Ю.Канрашина. Формальная модель пространства: структура и некоторые основные понятия.	59
Е.В.Падучева. Прагматические источники связности диалога.	61
Н.Н.Перцова. О некоторых закономерностях образования новых значений слов.	64
Э.В.Попов, А.Б.Преображенский. Определение функций и механизмов систем, взаимодействующих с многофункциональными базами данных.	66
Э.В.Попов, А.Б.Преображенский. От лингвистических процессоров – к пакету прикладных программ.	69
Е.Э.Разлогова. Проблема "точки зрения" в лексическом выводе.	72
Е.Н.Саввина. Об одной модели эллиптического и сочинительного сокращения в симметричных конструкциях.	75
М.Э.Садужеэр. Субъективная модальность как средство организации дискурса.	78
Г.В.Сенин. От анализа естественного языка к адаптивному диалогу.	80
И.Я.Сильдмяз. Формируются ли значения мыслями или мысли значениями?	83
Е.Г.Устинова. О конструкциях с числительными в русском языке.	85
В.Ф.Хорошевский. Реализация Л-процессоров в комплексе ПОСОХ методом конвертирования на входные языки СУБД.	87
А.С.Чехов. О единой семантической природе синтаксических и анафорических отношений.	90
З.М.Шаляпина. О понятии лингвистической нормативности в свете задач автоматической обработки текстов.	93

СЕМИНАР ПРОЕКТА "ДИАЛОГ".

Тезисы.

На русском языке.

Тартуский государственный университет.

ЭССР, 202400, г.Тарту, ул.Линксооли,18.

Ответственный редактор И. Силдмэе.

Корректор В. Бреков.

Сдано в печать 6.01. 1982.

МВ 02908.

Формат 60x84/16.

Бумага писчая.

Машиннопись. Ротапринт.

Условно-печатных листов 5,81.

Учетно-издательских листов 5,18.

Печатных листов 6,25.

Заказ № 5.

Тираж 200.

Цена 35 коп.

Типография ТГУ, ЭССР, 202400, г.Тарту, ул.Пялсона, 14.