

УДК 004.9, 004.89, 001.2

**КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА:
ВОЗНИКНОВЕНИЕ, ОСНОВНЫЕ НАПРАВЛЕНИЯ И ОСОБЕННОСТИ**

**COMPUTATIONAL LINGUISTICS:
APPEARANCE, MAIN DIRECTIONS AND FEATURES**

©**Калугян К. Х.**

канд. экон. наук,

Ростовский государственный
экономический университет (РИНХ)»

г. Ростов-на-Дону, Россия, kalugyan@yandex.ru

©**Kalugyan K.**

Ph.D.

Rostov State University of Economics (RSUE)

Rostov-on-Don, Russia, kalugyan@yandex.ru

©**Лозина Е. Н.**

Ростовский государственный
экономический университет (РИНХ)»

г. Ростов-на-Дону, Россия, westa78@mail.ru

©**Lozina E.**

Rostov State University of Economics (RSUE)

Rostov-on-Don, Russia, westa78@mail.ru

Аннотация. Рассматривается компьютерная лингвистика как молодое научное направление, причины и история ее возникновения, основные направления компьютерной лингвистики.

Abstract. Computational linguistics as the young scientific direction, the reasons and history of her appearance, the main directions of computational linguistics are considered.

Ключевые слова: компьютерная лингвистика, Джорджтаунский проект, автоматический анализ текстов, синтез текстов, программа-переводчик, модели коммуникации.

Keywords: computational linguistics, Georgetown project, automatic analysis of texts, synthesis of texts, translation program, communication models.

Лингвистика (языкознание, языковедение) как наука существует давно, для нее определены задачи, есть и развиваются основные направления, как в теоретическом, так и в практическом плане, например, терминоведение, лексикография, лингводидактика, переводоведение.

Компьютерная лингвистика – молодое и быстро развивающееся научное направление, возникшее на стыке лингвистики, вычислительной техники и математики. Условием ее появления стали две основные причины. С одной стороны, ученые-лингвисты предположили (и не ошиблись), что математика и информатика помогут обрести лингвистике недостающую ей точность. Появлялась возможность автоматизировать такие процессы, как анализ и синтез текстов, их обработка, ведение лексических и словарных картотек, и др. С другой стороны, с появлением компьютеров возникла проблема общения с ними непрофессиональных пользователей, самой удобной формой для которых является естественный язык. Поэтому для новой науки определилось свое поле деятельности – изучение и понимание законов и особенностей использования естественного языка в процессе общения людей между собой с

последующей задачей переноса этих законов на процесс общения человека с компьютером [1]. Т.е., главная цель компьютерной лингвистики – использование математических моделей, алгоритмов и программ для описания и обработки естественных языков и любой языковой информации.

Исследования в области компьютерной лингвистики начались с середины 20 в., а именно в 1947-48 гг., когда в Лондонском университете впервые попытались создать автоматический словарь. Самым важным событием с исторической точки зрения здесь явился 1954 г. Ученые Института языка и лингвистики Джорджтаунского университета и программисты IBM выдвинули идею (которая в дальнейшем получила название Джорджтаунский проект или Джорджтаунский эксперимент): используя ЭВМ, автоматизировать процесс перевода текстов с одного языка на другой. Исследования в этой области в нашей стране также начались в 1954 г. [1].

Основными направлениями компьютерной лингвистики являются [1]:

- анализ текстов на естественном языке;
- синтез текстов на естественном языке;
- понимание текстов;
- оживление тестов;
- модели коммуникации.

Анализ текстов на естественном языке. С появлением компьютеров исследования лингвистов по изучению текста в целом и каждого предложения в частности продолжились и получили новое направление.

Для автоматического анализа текстов необходимо пройти несколько последовательных этапов:

- предварительный анализ,
- морфологический анализ,
- поверхностный синтаксический анализ,
- глубокий синтаксический анализ,
- поверхностный семантический анализ,
- глубокий семантический анализ,
- прагматический анализ,
- выявление текстовых структур.

Первый этап – подготовительный: из текста устраняются опечатки и ошибки, расшифровываются сокращения и т.п.

Второй этап – морфологический анализ. На этом этапе обрабатываются и анализируются отдельные слова: в них выделяются основы (корни слов) и изменяемые части слов (приставки, суффиксы, окончания). Корни слов используются для поиска значений слов из специального словаря, а изменяемые части слов – для установления грамматических отношений между словами в рамках одного предложения, которые выделяются на этапе поверхностного синтаксического анализа. Его результатом является поверхностная синтаксическая структура для каждого предложения текста. Ее можно представить в виде размеченного дерева – графа. Узлы дерева – это словоформы или группы словоформ, ветви дерева – синтаксические связи между словами в рамках одного предложения (Рисунок). В зависимости от конкретного предложения в такой структуре могут отсутствовать некоторые элементы.

На этапе глубокого синтаксического анализа строится глубинная синтаксическая структура, которая затем, на этапах семантического анализа, преобразуется в поверхностную и глубинную семантические структуры, которые позволяют определить смысловые связи между предложениями текста. Последующие этапы (прагматический анализ и выявление текстовых структур) позволяют строить при необходимости прагматические и текстовые

структуры, результатом которых могут быть план текста, его аннотация, автореферат.

Перечисленные этапы охватывают всю задачу анализа текстов на естественном языке. При анализе конкретного текста необходимость в исполнении какого-либо этапа зависит от целей, для которых он осуществлялся [1].

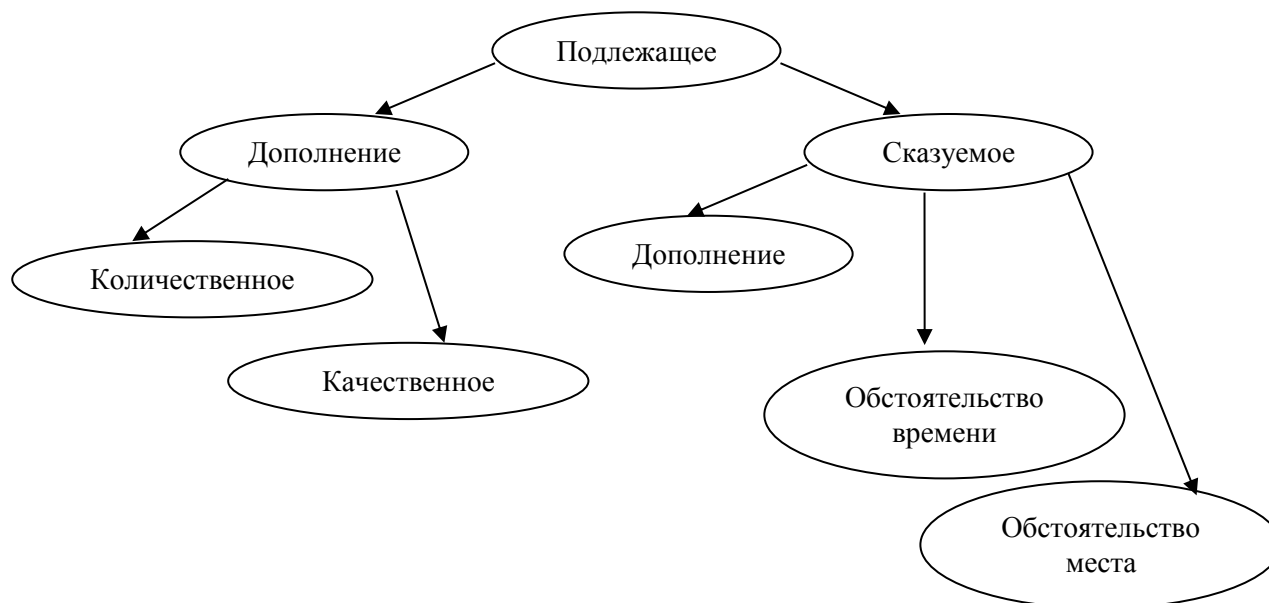


Рисунок. Поверхностная синтаксическая структура предложения (полный вариант представления)

Синтез текстов на естественном языке. Задача синтеза является противоположной по отношению к задаче анализа. Если заданы тема и цель будущего текста, то это значит, что определена прагматическая структура текста. Ее декомпозируют в прагматические структуры отдельных предложений, и для каждого из них все этапы анализа выполняются в обратном порядке. Например, если имеется план текста или его краткая аннотация, в процессе синтеза должен сформироваться полный вариант текста (существуют соответствующие компьютерные программы).

С анализом текстов мы сталкиваемся, хоть и не осознанно, но постоянно, работая в текстовом редакторе, используя программу-переводчик или электронный словарь. Когда в редакторе выполняется проверка правописания (орфографии, пунктуации, синтаксиса), то неявно для пользователей выполняются морфологический, синтаксический и семантический анализ текста. Или при переводе, чтобы распознать текст на исходном языке, выполняется его анализ, а для получения текста перевода выполняется его синтез, опирающийся на результаты уже проведенного анализа [2].

Понимание текстов. Проблема понимания текстов на естественном языке включает не только лингвистические аспекты. С ней тесно связаны задачи, которые решаются в рамках психологии, философии и семиотики.

Оживление текста. Самое молодое направление в компьютерной лингвистике. У людей зрительные представления об объектах и ситуациях и готовность описать увиденное с помощью родного языка всегда взаимосвязаны. Т.е. текст и сопутствующая ему зрительная картинка оказываются объединенными в сознании человека и интегрированными в некоторое единство. И задача, стоящая сейчас перед специалистами в области компьютерной лингвистики, – изучить, как происходит эта интеграция, и как по одной составляющей представления появляется вторая.

Модели коммуникации. Развитие компьютерных технологий способствовало появлению искусственных систем, которые могли воспринимать и понимать человеческую речь и тексты на естественном языке. Это, в свою очередь, создало предпосылки для непосредственного общения человека и компьютера и, как следствие, повысило интерес лингвистов к процессам, сопутствующим организации и ведению диалога. Примерами моделей коммуникации могут быть [1]:

построение сценария диалога на основе целей, которые ставит перед собой активная сторона; поддержка выбранного сценария с учетом интересов партнера; возможное противодействие партнера используемому сценарию; маскировка истинных намерений говорящего; организация пассивной поддержки коммуникационного процесса и др.

Список литературы

1. Информатика: энциклопедический словарь для начинающих / Сост. Д. А. Поспелов. М.: Педагогика-Пресс, 1994. 352 с.
2. Калугян К. Х. Информационные технологии в лингвистике // Информационные системы, экономика, управление трудом и производством: Ученые записки. Вып. 12. Ростов-на-Дону: РГЭУ «РИНХ», 2008. с. 73-76.

References:

1. Informatika: entsiklopedicheskii slovar' dlya nachinayushchikh / Sost. D. A. Pospelov. M.: Pedagogika-Press, 1994. 352 p.
2. Kalugyan K. Kh. Informatsionnye tekhnologii v lingvistike // Informatsionnye sistemy, ekonomika, upravlenie trudom i proizvodstvom: Uchenye zapiski. Vyp. 12. Rostov-na-Donu: RGEU "RINKh", 2008. P. 73-76