



# КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Авторы: А. Н. Баранов

КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА (калька с англ. computational linguistics), одно из направлений [прикладной лингвистики](#), в котором для исследования языка и моделирования функционирования языка в тех или иных условиях, ситуациях и проблемных сферах разрабатываются и используются компьютерные программы, компьютерные технологии организации и обработки данных. С др. стороны, это область применения компьютерных моделей языка в лингвистике и смежных с ней дисциплинах. Как особое науч. направление К. л. оформилась в европ. исследованиях в 1960-х гг. Поскольку англ. прилагательное computational может переводиться и как «вычислительный», в лит-ре встречается также термин «вычислительная лингвистика», однако в отеч. науке он приобретает более узкое значение, приближающееся к понятию «квантитативная лингвистика».

Часто к К. л. относят термин «квантитативная лингвистика», который характеризует междисциплинарное направление в прикладных исследованиях, где в качестве осн. инструмента изучения языка и речи используются количественные или статистич. методы анализа. Иногда квантитативная (или количественная) лингвистика противопоставляется комбинаторной лингвистике. В последней доминирующую роль занимает «неколичественный» математич. аппарат – теория множеств, математич. логика, теория алгоритмов и т. д. С теоретич. точки зрения использование статистич. методов в языкознании позволяет дополнить структурную модель языка вероятностным компонентом, т. е. создать теоретич. структурно-вероятностную модель, обладающую значит. объяснительным потенциалом. В прикладной области квантитативная лингвистика представлена прежде всего использованием фрагментов этой модели, используемых для лингвистич. мониторинга функционирования языка, дешифровки кодированного текста, авторизации/атрибуции текста и т. п.

Термин «К. л.» и проблематика этого направления часто связываются с моделированием общения, и прежде всего – с обеспечением взаимодействия человека с ЭВМ на естественном или ограниченном естественном языке (для этого создаются спец. системы обработки естеств. языка), а также с теорией и практикой информационно-поисковых систем (ИПС). Обеспечение общения человека с ЭВМ на естеств. языке иногда обозначается термином «обработка естественного языка» (перевод с англ. яз. термина Natural Language Processing). Это направление К. л. возникло в кон. 1960-х гг. за рубежом и развивалось в рамках науч.-технологич. дисциплины, именуемой искусств. интеллектом (работы Р. Шенка, М. Лебовица, Т. Винограда и др.). По своему смыслу словосочетание «обработка естественного языка» должно охватывать все области, в которых компьютеры используются для обработки языковых данных. На практике, однако, закрепилось более узкое понимание термина – разработка методов, технологий и конкретных систем, обеспечивающих общение человека с ЭВМ на естеств. или ограниченном естеств. языке.

К К. л. в определённой степени могут быть отнесены работы в области создания гипертекстовых систем, рассматриваемых как особый способ организации текста и даже как принципиально новый вид текста, противопоставленный по мн. своим свойствам обычному тексту, сформированному в гутенберговской традиции

книгопечатания (см. [Гутенберг](#)).

К компетенции К. л. относится и [автоматический перевод](#).

В рамках К. л. возникло и сравнительно новое, активно развивающееся с 1980–90-х гг. направление – корпусная лингвистика, где разрабатываются общие принципы построения лингвистич. корпусов данных (в частности, корпусов текстов) с использованием совр. компьютерных технологий. Корпуса текстов – это коллекции специально подобранных текстов книг, журналов, газет и т. д., перенесённые на машинные носители и предназначенные для автоматич. обработки. Один из первых корпусов текстов был создан для амер. варианта англ. языка в Брауновском ун-те (т. н. Брауновский корпус) в 1962–63 под рук. У. Френсиса. В России с нач. 2000-х гг. в Ин-те рус. языка им. В. В. Виноградова РАН разрабатывается Нац. корпус рус. языка, состоящий из представительной выборки русскоязычных текстов объёмом порядка 100 млн. словоупотреблений. Кроме собственно конструирования корпусов данных, корпусная лингвистика занимается созданием компьютерных инструментов (компьютерных программ), предназначенных для извлечения разнообразной информации из текстовых корпусов. С точки зрения пользователя, к корпусам текстов предъявляются требования представительности (репрезентативности), полноты и экономичности.

К. л. активно развивается и в России, и за рубежом. Поток публикаций в этой области очень велик. Кроме тематич. сб-ков, в США с 1984 ежеквартально выходит ж. «Computational Linguistics» («Компьютерная лингвистика»). Большую организац. и науч. работу проводит Ассоциация по компьютерной лингвистике (The Association for Computational Linguistics), которая имеет региональные структуры по всему миру (в частности, европ. отделение). Каждые два года проходят междунар. конференции КОЛИНТ (в 2008 конференция проходила в Манчестере). Осн. направления К. л. обсуждаются также на ежегодной междунар. конференции «Диалог», организуемой Рос. НИИ искусственного интеллекта, филологич. ф-том МГУ, Яндексом и рядом др. организаций. Соответствующая проблематика широко представлена также на междунар. конференциях по искусств. интеллекту разных уровней.

## Литература

Лит.: Звегинцев В. А. Теоретическая и прикладная лингвистика. М., 1968; Пиотровский Р. Г., Бектаев К. Б., Пиотровская А. А. Математическая лингвистика. М., 1977; Городецкий Б. Ю. Актуальные проблемы прикладной лингвистики // Новое в зарубежной лингвистике. М., 1983. Вып. 12; Кибрик А. Е. Прикладная лингвистика // Кибрик А. Е. Очерки по общим и прикладным вопросам языкознания. М., 1992; Kennedy G. An introduction to corpus linguistics. L., 1998; Bolshakov I. A., Gelbukh A. Computational linguistics: models, resources, applications. Mex., 2004; Национальный корпус русского языка: 2003–2005. М., 2005; Баранов А. Н. Введение в прикладную лингвистику. 3-е изд. М., 2007; Компьютерная лингвистика и интеллектуальные технологии. М., 2008. Вып. 7.