

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ЛИНГВИСТИКЕ

Нгуен Минь Хиен

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Петрова Н.Е. – к.филол.н., доцент

Статья посвящена исследованию использования программного обеспечения информационных технологий в лингвистике. Дается определение слову лингвистика. Обращается внимание на отношение между лингвистической наукой и информационными технологиями в настоящее время. Приводятся примеры аппаратных, программных и лингвистических средств. Описываются области применения информационных технологий в лингвистике. Называются прикладные разделы компьютерной лингвистики.

Информационные технологии позволяют активизировать и эффективно использовать информационные ресурсы общества, которые сегодня являются наиболее важным стратегическим фактором его развития. Опыт показывает, что активизация, распространение и эффективное использование информационных ресурсов (научных знаний, открытий, изобретений, технологий, передового опыта) позволяют получить существенную экономию других видов ресурсов: сырья, энергии, полезных ископаемых, материалов и оборудования, людских ресурсов, социального времени [1]. Информационные технологии играют в настоящее время ключевую роль также и в процессах получения и накопления новых знаний. При этом, на смену традиционным методам информационной поддержки научных исследований путем накопления, классификации и распространения научно-технической информации приходят новые методы, основанные на использовании вновь открывающихся возможностей информационной поддержки фундаментальной и прикладной науки, которые предоставляют современные информационные технологии.

Лингвистика (от лат. *lingua* «язык»), языкознание, языковедение – наука, изучающая языки. Это наука о естественном человеческом языке вообще и обо всех языках мира как его индивидуализированных представителях [2].

В широком смысле слова лингвистика подразделяется на научную и практическую. Чаще всего под лингвистикой подразумевается именно научная лингвистика. Лингвистика связана с семиотикой как наукой о знаковых системах. Лингвистикой профессионально занимаются учёные-лингвисты. Лингвистика изучает не только существующие (существовавшие или возможные в будущем) языки, но и человеческий язык вообще. Язык не дан лингвисту в прямом наблюдении; непосредственно наблюдаемы лишь факты речи, или языковые явления, то есть речевые акты носителей живого языка вместе с их результатами (текстами) или языковой материал (ограниченное число письменных текстов на мёртвом языке, которым уже никто не пользуется в качестве основного средства общения) [3].

Информационные технологии помогают развивать теоретическую подготовку лингвистов, а также закреплять полученные знания, готовиться к жизни в информационном пространстве, достигать максимального качества в профессиональной деятельности. Получать и накапливать знания помогают методы информационного моделирования, искусственный интеллект, когнитивная компьютерная графика [4].

В настоящее время лингвистическая наука находится в тесной связи с информационными технологиями. С помощью технологий можно успешно решать всевозможные задачи лингвистики. В то же время сама эта наука оказывает значительное воздействие на формирование современных инноваций.

Лингвистика и новые информационные технологии не стоят на месте. Перспективным направлением в настоящее время является компьютерная лингвистика. С её помощью можно создавать мощные словари, изучать их семантико-синтаксическую структуру, создавать базовые процедуры концептуального, семантико-синтаксического и морфологического анализа и синтеза текстов. Среди важнейших инструментов компьютерной лингвистики можно назвать формальную логику, грамматику и алгоритмы [5].

С наукой лингвистикой тесно связано приобретение новых знаний. С развитием информационных технологий меняются концептуальные представления, технические средства, методы и сферы применения, а также информационная среда в целом. Пройдя определенные эволюционные этапы и вступив в эпоху технического прогресса, лингвистика прочно компьютеризировалась [6].

Для выполнения объёмных расчётов над лингвистическими данными, а также для лингвистического моделирования удобно использовать электронные вычислительные машины (или компьютеры).

В зависимости от назначения программных средств различают системное и прикладное программное обеспечение. Системные программы служат управлению работой аппаратных средств и включают операционные системы, утилиты, драйверы и некоторые другие виды программ. Прикладные программы предназначены для конечного пользователя и позволяют ему выполнять различные операции над информацией: создавать и обрабатывать текст (текстовые редакторы), обрабатывать графические изображения (графические редакторы), работать над звуковой и видеoinформацией (мультимедийные программы), создавать электронные таблицы для обработки статистических данных (электронные таблицы) и т.д. Для лингвиста особенно полезными являются такие виды прикладных программ, как электронные переводчики и словари, а также мультимедийные обучающие программы.

Наряду с аппаратным и программным обеспечением (ПО) информационных технологий некоторые исследователи используют также понятие *lingware* (или *linguware*), которым обозначаются все лингвистические компьютерные ресурсы (грамматические справочники, словари, энциклопедии, лингвистические базы данных и т.п.) [7, с. 22].

Совокупность аппаратных, программных и лингвистических средств, необходимых для автоматической обработки лингвистических данных, обозначается понятие автоматическое рабочее место (АРМ) лингвиста. АРМ лингвиста будет включать сам компьютер, операционное и базовое прикладное ПО, а также всевозможные лингвистические компьютерные ресурсы, касающиеся родного и изучаемых иностранных языков [7, с. 22].

В зависимости от специализации АРМ лингвиста может дополняться прикладными программами и лингвистическими ресурсами, связанными с переводом или обучением иностранному языку. Задачей обучающихся является постоянная актуализация своего АРМ, включающая поддержание современного состояния аппаратного и программного обеспечения, а также постоянное пополнение собственной лингвистической ресурсной базы, т.е. поиск, сохранение, приобретение или создание лингвистических справочников, словарей и баз данных [7, с. 22].

Информационные технологии применяются в лингвистике в качестве автоматического анализа и синтеза звучащей речи для преобразования печатного текста существующего в цифровой форме, в звучащий текст на естественном человеческом языке и обратно. Процесс автоматического анализа речи включает ввод слов в компьютер через микрофон, начитанных разными дикторами, их спектральную обработку и создание набора признаков, своеобразного образца слова, который выступает знаком языка. При распознавании звучащей речи реальные признаки составляющих ее единиц сравниваются с признаками и образцами слов, существующими в памяти машины. Результатом сравнения является транскрипция или орфографическая запись слова [7, с. 27].

Информационные технологии применяются в лингвистике также в автоматическом распознавании текста. Для ввода информации в компьютер используются специальные устройства – клавиатура, мышь и др., но наиболее удобным инструментом для ввода большого количества печатных текстов является сканер. Программа автоматического распознавания текста (OCR-программа) – это компьютерная программа, позволяющая преобразовать текст с бумажного носителя в электронный текстовый файл, который в дальнейшем может обрабатываться человеком в любом текстовом редакторе. В целом точность распознавания OCR программ на текстах хорошего и среднего качества достигает 99%, что позволяет считать проблему массового ввода печатных текстов в компьютер практически решенной [7, с. 35].

Информационные технологии используются в лингвистике и качестве автоматического аннотирования и реферирования текста. Некоторые исследователи считают реферат и аннотацию синонимами, а некоторые предлагают разводить эти понятия, определяя аннотацию как краткое изложение содержания документа, дающее общее представление о его теме. Согласно этому определению в отличие от реферата, знакомящего читателя с сутью излагаемого в документе содержания, аннотация выполняет лишь сигнальную функцию (есть публикация на определенную тему). Наиболее простыми системами автоматического реферирования и аннотирования является функция AutoSummarize в MS Word, системы Intelligent Text Miner, Oracle Context и Inxight Summarizer (компонент поискового механизма AltaVista) (IBM). Правда, возможности этих программ ограничены выбором оригинальных фрагментов из исходного документа и их соединением в короткий текст. В целом можно констатировать, что автоматические рефераты и аннотации представляют собой, по сути, квазирефераты, т.е. результатом автоматической компрессии текста в большинстве случаев становится либо

набор ключевых слов, либо перечень ключевых предложений, что, впрочем, в значительной степени помогает решить задачу аннотирования и реферирования большого объема текстов в малые сроки [7, с. 40].

В лингвистике информационные технологии используются также и в автоматическом анализе и синтезе текста. Автоматический анализ и синтез текста являются важными задачами компьютерной лингвистики как с точки зрения развития теории (разработки лингвистических основ создания искусственного интеллекта), так и с точки зрения реализации практических нужд человека, например, создания эффективных систем машинного перевода. На современном этапе системы компьютерного моделирования письменных диалогов на естественном языке используют более сложные алгоритмы. В частности, создан специальный язык разметки для искусственного интеллекта AIML (Artificial Intelligence Markup Language), используемый для создания виртуальных агентов (или ботов). Боты, моделирующие диалог с собеседником, используются в компьютерных играх и на корпоративных веб-страницах, например, для ответов на вопросы пользователя о возможностях мобильного оператора или торговой сети [7, с. 44].

Компьютерная лингвистика обладает прикладными разделами. Корпусная лингвистика – раздел прикладной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов при помощи компьютеров. Центральное понятие корпусной лингвистики – лингвистический корпус – определяется как совокупность специально отобранных текстов, размеченных по различным лингвистическим параметрам и обеспеченных системой поиска [7, с. 57].

Если говорить о использовании информационных технологии в лингвистике, то необходимо остановиться на понятии компьютерные лексикографий. Компьютерная лексикография представляет собой раздел прикладной лингвистики, нацеленный на создание компьютерных словарей, лингвистических баз данных и разработку программ поддержки лексикографических работ. Автоматические словари такого типа практически повторяют структуру словарной статьи обычных словарей, однако они обладают функциями, недоступными своим прототипам, например, осуществляют сортировку данных по полям словарной статьи (ср. отбор всех прилагательных), проводят автоматический поиск всех вокабул, имеющих в толковании определенный семантический компонент, и т.д. В целом констатируем, что компьютерная лексикография, направленная на создание электронных словарей, представляет собой весьма перспективное и нужное направление компьютерной лингвистики, поскольку создаваемые ею продукты – электронные словари – отличаются многогранностью, мультимедийностью, интеграцией новейших технологических решений, актуальностью материала и отвечают потребностям пользователя в организации доступа к необходимой информации [7, с. 65].

Компьютерная терминография также связана с использованием информационных технологии в лингвистике. Она занимается построением специальных терминологических словарей занимается терминография, представляющая собой особый раздел лексикографии. В то же время терминография тесно связана с терминоведением – наукой о терминах. Соответственно, компьютерная терминография – это наука о составлении электронных терминологических словарей. При анализе составляющих словарной статьи терминологического словаря можно заметить, что такой словарь требует еще более тщательной работы, чем обычный словарь [7, с. 76].

Информационные технологии активно используются в лингвистике в целях осуществлении машинного перевода. Вопросы машинного перевода составляют одну из центральных областей использования информационных технологий в лингвистике. Это обусловлено не только тем, что в машинном переводе как в фокусе концентрируются все проблемы компьютерной лингвистики от способов анализа содержания до синтеза словоформы, предложения и целого текста, но и постоянно возрастающей практической потребностью современного общества в переводе значительного количества текстов различной функциональной направленности. машинный перевод, представляющий собой процесс передачи содержания текста на одном языке средствами другого языка с использованием компьютеров, является одним из первых и не теряющих своей актуальности направлений компьютерной лингвистики. Процесс машинного перевода может предполагать разную степень активности человека в его выполнении, что обуславливает многообразие его форм, выбор которых зависит от целей перевода и его условий [7, с. 81].

Компьютерное обучение языкам. Одной из важных практических областей применения компьютеров в лингвистике является компьютерное обучение языкам (Computer Assisted Language Learning, CALL). Компьютеризация и информатизация являются характерными особенностями современного обучения в целом, поскольку применение современных информационно-коммуникационных технологий в обучении позволяет сделать его более эффективным, повысить мотивацию обучающихся и сократить затраты человеческого труда. Кроме того, применение компьютеров в полной мере соответствует другим современным тенденциям образования: его деятельности и личностно-ориентированному характеру [7, с. 91].

Информационно-поисковые системы. В современном мире, который буквально пронизан постоянно нарастающими объемами информации, для человека, использующего эту информацию с целью ее превращения в знания, встает проблема ориентации. Чтобы не захлебнуться в информационном потоке, нам нужны техники отбора, фильтрации и оценки. Упорядоченная совокупность документов и информационных технологий, предназначенных для хранения и поиска информации, представленной в виде текстов или их частей (фактов), получила название информационно-поисковой системы. Итак, информация не просто дается человеку «на кончиках пальцев», а предполагает сложные и трудоемкие процессы сортировки и отбора. С

этими задачами в значительной степени помогают справиться современные автоматические информационно-поисковые системы, в частности поисковые системы Всемирной паутины [7, с. 98].

Таким образом, информационные технологии в лингвистике – это совокупность законов, способов и средств получения, хранения, передачи, распространения, преобразования информации о языке и законах его функционирования с помощью компьютеров. Информационные технологии в настоящее время являются неотъемлемой частью любой сферы профессиональной деятельности и играют очень важную роль в лингвистике, помогают эффективно решать различные проблемы в профессиональной деятельности. Применение информационных технологий в лингвистике сегодня является необходимостью в современном обществе.

Список использованных источников:

1. Информационная технология [Электронный ресурс]. – Режим доступа: <https://studfile.net/preview/2484532/page:2/>. – Дата доступа: 04.11.2022.
2. Лингвистика [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org/wiki/Лингвистика>. – Дата доступа: 04.11.2022.
3. Лингвистика [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org/wiki/Лингвистика>. – Дата доступа: 04.11.2022.
4. Информационные технологии в лингвистике [Электронный ресурс]. – Режим доступа: <https://qwizz.ru/информационные-технологии-лингвис/?ysclid=laldn7tfu4899232737>. – Дата доступа: 04.11.2022.
5. Информационные технологии в лингвистике [Электронный ресурс]. – Режим доступа: <https://qwizz.ru/информационные-технологии-лингвис/?ysclid=laldn7tfu4899232737>. – Дата доступа: 04.11.2022.
6. Информационные технологии в лингвистике [Электронный ресурс]. – Режим доступа: <https://qwizz.ru/информационные-технологии-лингвис/?ysclid=laldn7tfu4899232737>. – Дата доступа: 04.11.2022.
7. Щипицина, Л. Ю. Информационные технологии в лингвистике : учеб. пособие / Л. Ю. Щипицина. – М. : ФЛИНТА Наука, 2013. – 128 с.