

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЛИНГВИСТИКЕ

Литвиненко И.А.

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

Петрова Н.Е. – к.филол.н., доцент

Статья посвящена использованию информационных технологий в лингвистике. Рассмотрены главные понятия из сферы лингвистики, связанные с использованием вычислительной техники. Называется основная задача прикладной лингвистики и её связь с современными информационными технологиями. Приведены примеры использования информационных технологий в современной лингвистике и возникающие при этом проблемы. Основная цель работы – показать важность применения методов информационных технологий в развитии современной лингвистики.

Информационные технологии в современном мире являются неотъемлемой частью любой сферы профессиональной деятельности, в том числе лингвистики. Если когда-то использование компьютеров и соответствующих программ в лингвистических исследованиях, переводе и в обучении языку не являлось обязательным, то сегодня будущим преподавателям иностранных языков, переводчикам и лингвистам необходимы знания, связанные с использованием информационных технологий в сфере своей профессиональной деятельности. В современном мире владение информационными технологиями ставится наряду с такими качествами, как умение читать и писать. В наши дни специалист с высшим образованием должен свободно ориентироваться в огромном информационном пространстве, иметь необходимые знания и навыки, в том числе: поиска, обработки и хранения информации с использованием новейших информационных технологий, компьютерных систем и сетей.

Лингвистика – это наука о естественном человеческом языке вообще и обо всех языках мира как индивидуальных его представителях. В широком смысле она занимается познанием языка и передачей результатов этого познания для их практических целей. Лингвистика делится на теоретическую и прикладную [1].

Прикладная лингвистика – это направление в языкознании, занимающееся разработкой методов решения практических задач, связанных с использованием языка. К задачам прикладной лингвистики относят те, появление которых обусловлено современной научно-технической революцией, характеризующейся укреплением взаимосвязи общественных, естественных и технических наук [2]. В русскоязычном мире начиная с 1950-х годов, прикладная лингвистика ассоциируется с компьютерными, информационными технологиями и автоматическими системами обработки информации [3].

Одним из основных назначений языка является его использование для передачи информации между людьми. Информация, кодируется с помощью языка. Уже в середине XX в. появляются специальные науки, изучающие информацию: информатика и кибернетика. Компьютеры в информационном обмене становятся средством кодирования, обработки, хранения и передачи больших массивов символьной информации. Совокупность законов, методов и средств получения, хранения, передачи, распространения и преобразования информации с помощью компьютеров получило обозначение «информационные технологии» [4]. При сужении этого понятия для его использования в особой профессиональной сфере (лингвистика) получаем сочетание «информационные технологии в лингвистике», понимаемое как совокупность законов, методов и средств получения, хранения, передачи, распространения и преобразования информации о языке и законах его функционирования с помощью компьютеров [5].

Проблемы языковой коммуникации «человек – компьютер – человек» и моделирования языка лежат в области исследований такой молодой науки как компьютерная лингвистика, которая образовалась на стыке информатики и лингвистики с началом внедрения первых вычислительных машин. Данное направление стало активно разрабатываться в 60–70 гг., и под ним в первую очередь понималось использование статических методов в языкознании. В конце XX века в связи с развитием компьютерных технологий и их активным применением в лингвистических задачах эта наука получила более четкое наименование «компьютерная лингвистика» [6].

Понятие «информационные технологии» в лингвистике относится в основном к задачам прикладной лингвистики. К их числу можно отнести, создание систем искусственного интеллекта, создание систем автоматического перевода, создание систем автоматического аннотирования и реферирования текстов, создание систем порождения текстов, создание систем обучения языку, создание систем понимания устной речи, создание систем генерации речи, создание автоматизированных информационно-поисковых систем, создание систем атрибуции и дешифровки анонимных и псевдоанонимных текстов, разработка различных баз данных (словарей, карточек, каталогов, реестров и т.п.) для гуманитарных наук, разработка различного типа автоматических словарей, разработка систем передачи информации в сети Интернет и т.д. Эти комплексные задачи включают целый ряд более мелких проблем. К их числу относится автоматизация следующих процессов: построение словарей текстов, морфологический анализ слова, определение значения

многозначного слова, синтаксический анализ предложения, поиск слова в словаре, порождение предложения и т.д. [7].

Потребность в создании систем автоматической обработки языка возникла, поскольку невозможно обучить всех пользователей программированию и оптимальной формой диалога человека и компьютера является диалог на естественном языке. Под обработкой устной речи понимается разработка методов, технологий и конкретных систем, которые обеспечивают общение человека на естественном или ограниченно естественном языке. Речевой диалог обладает преимуществами: устное общение не требует специальной подготовки пользователя, диалог освобождает руки и зрение, за счет системы распознавания голоса возможна защита от нежелательного доступа к объекту, возможность использования людьми с ограниченными возможностями. Однако такая связь имеет и недостатки: подверженность шумовым помехам, невозможность неограниченного ввода данных в компьютер в течение длительного времени. Такие системы необходимы для организации приема и озвучивания сообщений, например, получение электронной почты по телефону, а также для перевода звучащей речи в текст в электронной форме. Компьютеры могут оказывать помощь и при обучении иностранному языку с помощью автоматических фонетических тренажеров [8].

Автоматическая обработка письменного текста применяется для распознавания текста, анализа и синтеза. Такие технологии необходимы для облегчения ввода печатного текста путем сканирования. Они базируются на статических данных о частоте употребления графем и на данных о возможных правилах сочетаемости графем в конкретном языке, а также о частотности определенных буквосочетаний. При автоматическом анализе текст последовательно преобразуется в его лексемно-морфологические, синтаксические и семантические представления, понятные компьютеру. Обратный процесс преобразования лексемно-морфологических, синтаксических и семантических компьютерных представлений в текст на естественном языке называется автоматическим синтезом 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 [9]. На сегодняшний день много вопросов вызывает возможность синтеза прагматической структуры текста из тех целей, которые стимулируют создание текста. Например, неясно, как структуру разбить на структуры предложений и как от них перейти к глубинным семантическим структурам.

Поиск информации является часто возникающей задачей в наши дни. Сейчас возникает необходимость в систематизации информации и облегчению её поиска. Эту проблему решают информационно-поисковые системы. Сущность любого поиска заключается в отборе информации по запросу в базе данных, где хранится информация. При этом не важно, в каком виде представлена информация. Для большей эффективности применяют специальные программы, производящие лингвистическую обработку текстов с учетом морфологии и синтаксиса, ибо не всегда простое использования названия документа является оптимальным выбором поискового образа [10].

В условиях все возрастающего количества текстов в окружающем человека мире возникает проблема поиска нужного документа и ознакомления с его содержанием. В решении проблемы помогают составление рефератов и аннотаций полнотекстовых документов. Они дают читателю представление о содержании исходных документов и позволяют оценить степень необходимости обращения к полным текстам каждой работы. Кроме того, рефераты и аннотации акцентируют внимание читателя на новых сведениях, т.е. позволяют за небольшой промежуток времени узнать много новой информации. 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 [11].

Все эти направления и многие другие развиваются с каждым днём благодаря современным компьютерным технологиям. На сегодняшний момент существует множество систем и методов, которые помогают не только лингвистам, но и рядовым пользователям при взаимодействии с компьютером. Сейчас эти сервисы облегчают жизнь людям в их повседневной жизни, а уже завтра, возможно, будут облегчать существование машин с искусственным интеллектом в повседневном общении с людьми на естественном языке.

Список использованных источников:

1. Соснина, Е. П. Введение в прикладную лингвистику / Е. П. Соснина. – Ульяновск : УлГТУ, 2012. – 110 с.
2. Zubov, A. V. Информационные технологии в лингвистике / А. В. Zubov, И. И. Zubova. – Москва : Академия, 2004. – 208 с.
3. Баранов, А. Н. Введение в прикладную лингвистику: учеб. пособие / А. Н. Баранов. – 5-е изд. – Москва : Моск. Гос. ун-т им. М. В. Ломоносова, 2017. – 368 с.
4. Щипицина, Л.Ю. Информационные технологии в лингвистике: учеб. пособие / Л. Ю. Щипицина. – Москва.: ФЛИНТА, 2013. – 314 с.
5. Zubov, A. V. Информационные технологии в лингвистике / А. В. Zubov, И. И. Zubova. – Москва : Академия, 2004. – 208 с.
6. Соснина, Е. П. Введение в прикладную лингвистику / Е. П. Соснина. – Ульяновск : УлГТУ, 2012. – 110 с.
7. Zubov, A. V. Информационные технологии в лингвистике / А. В. Zubov, И. И. Zubova. – Москва : Академия, 2004. – 208 с.
8. Шилихина, К. М. Основы прикладной лингвистики: учеб. пособие / К. М. Шилихина. - Воронеж: Изд-во ВГУ, 2006. - 51 с.

56-я Научная Конференция Аспирантов, Магистрантов и Студентов БГУИР, Минск, 2020

9. Щипицина, Л.Ю. Информационные технологии в лингвистике: учеб. пособие / Л. Ю. Щипицина. – Москва.: ФЛИНТА, 2013.
10. Шилихина, К. М. Основы прикладной лингвистики: учеб. пособие / К. М. Шилихина. - Воронеж: Изд-во ВГУ, 2006. – 51 с.
11. Щипицина, Л.Ю. Информационные технологии в лингвистике: учеб. пособие / Л. Ю. Щипицина. – Москва.: ФЛИНТА, 2013.