**ЛИНГВИСТИКА И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ**

Существует достаточно большое число определений понятия **лингвистика.**

**В** «Лингвистическом энциклопедическом словаре» (1990)

**Лингвистика** (языкознание, языковедение) – наука о естественном человеческом языке вообще и о всех языках мира как индивидуальных его представителях*.*

– исследует сущность и природу языка, проблему его происхождения и общие законы его развития и функционирования (Ю. С. Маслов)*.*

Существуют различные подходы выделения внутри лингвистики отдельных ветвей или направлений*.*

*1 подход* (конец 60-х — начало 70-х годов XX века).Четко отделяются друг от друга теоретическая и прикладная лингвистика. Эти два направления лингвистики взаимосвязаны и дополняют друг друга, что успешное функционирование прикладной лингвистики возможно лишь на базе лингвистических теорий, разработанных в рамках теоретической лингвистики. Проблемами прикладной лингвистики являются машинный перевод, автоматическая обработка речевой информации, порождающая грамматика, дескриптивная лингвистика, математическая лингвистика и др.

*2 подход* Р.Г.Пиотровский. Четко разделяет теоретическую и прикладную лингвистику, но не признавая последнюю как самостоятельный раздел языкознания, выделяет структурное и математическое языкознание, а также «новую» лингвистику, к которой относит инженерную лингвистику*.* В последней он выделяет далее вычислительную лингвистику, экспериментальную фонетику, лингвистическое обеспечение систем научно-технической информации.

*3 подход* Коллектив авторов Ленинградского университетавыделил в языкознании три взаимосвязанных направления: теоретическую лингвистику, экспериментальную лингвистику и прикладную лингвистику. При этом структурную лингвистику авторы считают симбиозом таких наук, как теоретическое языкознание, психология, логика, семиотика, математика, математическая лингвистика. С их точки зрения, это совокупность методов теоретической лингвистики и математики*.* В последние годы в рамках прикладной лингвистики выделяют также компьютерную лингвистику.

Как видно из вышесказанного, «прикладная лингвистика» — понятие, до сих пор не имеющее четкого определения и конкретного конечного перечня решаемых ею задач.

Одно из первых определений понятия **прикладная лингвистика** принадлежит В. В. Звегинцеву. Он определил ее как новую область лингвистики, «которая осуществляет реализацию лингвистических знаний с целью решения всякого рода практических задач»*.*

В «Лингвистическом энциклопедическом словаре» прикладная лингвистика определяется как «направление в языкознании, занимающееся разработкой методов решения практических задач, связанных с использованием языка»*.* При этом задачи прикладной лингвистики делятся на традиционные, или «вечные», и «новые»*.* К числу первых относят создание и совершенствование письменности, разработку систем транскрипции устной речи, систем транслитерации иноязычных слов, унификацию и стандартизацию научно-технической терминологии, создание словарей различных типов, перевод с языка на язык, обучение языку и т.д.К новым задачам прикладной лингвистики относятся те, появление которых обусловлено современной научно-технической революцией, характеризующейся укреплением взаимосвязи общественных, естественных и технических наук.

Наиболее удачным представляется определение прикладной лингвистики и перечня решаемых ею задач, сделанное А. Е. Кибриком. По мнению автора, «прикладная лингвистика — раздел языкознания, в котором разрабатываются методы решения практических задач, связанных с оптимизацией использования языка как важнейшего средства человеческой коммуникации»*.* Множество задач, решаемых прикладной лингвистикой, А. Е. Кибрик выделяет с опорой на те функции языка, которые оптимизируются задачами прикладной лингвистики. В частности, он выделяет следующие четыре функции:

1. оптимизация способов фиксации и хранения речевой информации;
2. оптимизация способов передачи информации;
3. оптимизация интеллектуальных способностей человека, связанных с использованием языка;
4. оптимизация использования языка как средства массовой коммуникации.

В рамках каждого из этих разделов автор выделяет конкретные задачи. Их также можно условно разделить на «вечные» и «новые». К числу задач первого типа, отмечаемых А. Е. Кибриком, можно отнести создание алфавитов и письменностей, создание систем транскрипции и транслитерации, задачи сурдопедагогики и задачи использования языка в медицине, языковое планирование, языковое строительство, нормализацию языка и т.д.

«Новые» задачи — это такие проблемы, решение которых возможно с использованием современных информационных технологий.

Рассматривая определение понятия «информационные технологии» по отношению к лингвистике, можно сказать, что **информационные технологии в лингвистике** — это совокупность законов, методов и средств получения, хранения, передачи, распространения, преобразования информации о языке и законах его функционирования с помощью компьютеров. Если соотнести это определение с теми задачами, которые решает современная прикладная лингвистика,то можно отметить, что понятие «информационные технологии» в лингвистике относится в основном к задачам прикладной лингвистики. К их числу можно отнести:

1. создание систем искусственного интеллекта;
2. создание систем автоматического перевода;
3. создание систем автоматического аннотирования и реферирования текстов;
4. создание систем порождения текстов;
5. создание систем обучения языку;
6. создание систем понимания устной речи;
7. создание систем генерации речи;
8. создание автоматизированных информационно-поисковых систем;
9. создание систем атрибуции и дешифровки анонимных и псевдоанонимных текстов;
10. разработка различных баз данных (словарей, карточек, каталогов, реестров и т.п.) для гуманитарных наук;
11. разработка различного типа автоматических словарей;
12. 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:

1. построение словарей текстов;
2. морфологический анализ слова;
3. определение значения многозначного слова;
4. синтаксический анализ предложения;
5. поиск слова в словаре;
6. порождение предложения и т.д.

Философы, психологи и другие специалисты отмечают, что в будущем социально защищенным может считаться лишь тот человек, который способен гибко перестраивать направление и содержание своей деятельности в связи со сменой технологий или требований рынка. Чтобы подготовить такого человека, необходимо заменить традиционную технологию получения новых знаний более эффективной организацией познавательной деятельности обучаемых в ходе учебного процесса. Это можно сделать с использованием современных информационных технологий. Именно они могут продемонстрировать обучаемому тот факт, что любой информационный ресурс представляет реальную ценность лишь в том случае, когда к нему организован соответствующий доступ. С их помощью будущих специалистов можно научить правильной организации хранения информации и выбору адекватных форм ее представления*.* «Интеллектуальная собственность, представленная в цифровом формате, станет главной "валютой" XXI века»*.*

Технология получения и распространения новых знаний уже сейчас неотделима от Internet. Формируется инфраструктура, объединяющая в единое целое глобальные и местные телекоммуникационные каналы, радио, телевидение, телефонные линии связи (одна планета — одна сеть). Все это не просто создает сверхинтеллект, а формирует новые измерения сознания, феномен сверхпсихологических изменений в личности человека*.* Широкое распространение в недалеком будущем получат видеоконференции и дистанционное обучение*.*

К концу XXI века ученые изучат принципы работы мозга на уровне отдельных нейронов и научатся обращаться с ним, как со сложным электронным объектом*.* Это даст новый толчок к созданию систем искусственного интеллекта, систем автоматического порождения текстов, их перевода, реферирования и т.д. В ближайшие годы должна найти решение проблема распознавания и синтеза устной речи, широкое распространение получат электронная коммерция (e-commerce) мобильная цифровая телефонная связь, генная инженерия*.*

В составе современных информационных технологий можно выделить следующие составляющие:

1. теоретические основы информационных технологий;
2. методы решения задач информационными технологиями;
3. средства решения задач, используемые в информационных технологиях:

а) аппаратные средства;

б) программные средства.

Рассмотрим подробнее эти составляющие.

Теоретическую основу информационных технологий составляют важнейшие понятия и законы информатики. В свою очередь понятие «информатика» тесно связано с понятием «информация».

Слово **информация** (от лат. *informatio* — 'разъяснение, изложение') в обычном житейском понимании обозначает некоторые сведения о внешнем и внутреннем мире, которые мы используем для регулирования своего поведения. Более строго это понятие раскрывается разными способами*.* Выберем один из них, который более всего подходит к рассматриваемым в лингвистическим задачам, и определим **информацию**

– как определенным образом связанные сведения, данные, понятия, отраженные в нашем сознании и изменяющие наши представления о реальном мире*.*

Информация обладает разными свойствами*.* Наиболее важными из них являются: ценность, достоверность, полнота, актуальность, логичность, компактность. ***Ценность*** информации определяется тем, насколько она важна для достижения цели, стоящей перед ее получателем. ***Полнота*** информации связана с тем, насколько много в ней сведений, позволяющих получателю информации достичь своей цели. ***Актуальность*** информации определяется необходимостью ее немедленного использования для достижения какой-либо цели. ***Компактность*** информации — способность представить ее в наиболее сжатом виде. Понятия ***достоверность*** и ***логичность*** информации не требуют особых пояснений.

Выделяют различные виды информации. При этом для ее классификации по видам разработано много подходов, использующих разнообразные признаки и особенности информации. Так, в зависимости от того, какими органами чувств воспринимается информация, ее делят на визуальную, аудиальную (звуковую, фонетическую), аудиовизуальную, тактильную. По направленности информации всем членам общества или каким-то его группам различают информацию массовую, предназначенную для всех членов общества, и специальную — для специалистов в различных областях науки, техники, культуры, производства. Специальную информацию подразделяют на научную, техническую, производственную, эстетическую и т.п.

В каждом виде специальной информации выделяют подвиды. Например, в зависимости от области науки и научной информации выделяют информацию физическую, математическую, биологическую, лингвистическую и т.д. Так, **лингвистической информацией** называют множество определенным образом связанных сведений, данных, понятий о языке и правилах его функционирования, отраженных в нашем сознании и влияющих на наше речевое поведение.

Слово «информатика» также не имеет единого определения. С современной точки зрения **информатика** — это наука о законах и методах получения, хранения, передачи, распространения, преобразования и использования информации в естественных и искусственных системах с применением компьютера.

В зависимости от вида информации выделяют различные типы информатики. Так, различают информатику социальную, экономическую, научную, научно-техническую, статистическую, биологическую, медицинскую и т.п.

Наука, изучающая законы и методы организации и переработки с помощью компьютера лингвистической информации, называется **лингвистической информатикой.** Вспомнив, что понимается под лингвистической информацией, можно сказать, что ***объектом исследования*** лингвистической информатики будет структура слов, словосочетаний, предложений, текстов.

Ее интересуют правила, объединяющие нижестоящие языковые единицы в вышестоящие, правила перевода предложений и текстов, способы построения рефератов и аннотаций, пути обучения языкам и целый ряд других вопросов, связанных с языком и речью.

Основным методом решения различных задач информационными технологиями является **метод моделирования.** Суть его заключается в том, что для решения какой-либо задачи строится модель некоторого объекта, явления или процесса. Этот метод используется человеком очень давно. Существуют различные определения понятий «модель» и «моделирование»*.*

**Модель** — это формализованное описание объекта, системы нескольких объектов, процесса или явления, выраженное конечным набором предложений какого-либо языка, математическими формулами, таблицами, графиками, специальными знаками или какими-нибудь схемами.

Описание считается формализованным, если оно понятно не только человеку, но и некоторому устройству, например компьютеру. Предположим, архитектор, разрабатывая план какого-либо города или поселка, строит его модель в виде таких специальных знаков, как квадрат, прямоугольник, круг, которые обозначают целые дома, заводы, улицы и т.п.

Модель взаимосвязи в треугольнике его сторон (катетов *а и b* гипотенузы с) выражается формулой *с2 = а2 + b2.*

Модель распределения словоформ какого-либо текста по частоте употребления может быть представлена в виде таблицы.

|  |  |
| --- | --- |
| Словоформа | Частота |
| Информация  Компьютер  Технология | 73  46  27 |

По отношению к моделируемому объекту, процессу или явлению модель должна удовлетворять целому ряду свойств. Важнейшими из них являются следующие*.*

1. Модель выступает в качестве упрощенного аналога изучаемого объекта (процесса, явления).
2. Модель не должна быть сложнее самого оригинала.
3. Метод изучения объекта (процесса, явления) путем его моделирования должен быть более экономным по сравнению с другими возможными методами изучения того же объекта.
4. Построенная модель должна быть предельно простой и логически корректной, не содержащей противоречий.
5. Модель должна по возможности иметь общий (универсальный) характер, позволяющий использовать ее для изучения других подобных объектов (процессов, явлений). Например, построив на материале английского текста модель его реферирования, опирающуюся на ключевые слова текста, необходимо, чтобы эта модель работала и для текстов других языков.

6. Модель должна отражать наиболее существенные черты реального объекта, процесса или явления, которые важны для проводимого в данный момент процесса моделирования.

Существуют различные виды моделей. При использовании информационных технологий в лингвистике выделяют следующие типы моделей*.*

1. ***Структурные модели*** служат для изучения и описания внутреннего строения некоторого объекта. Например, такая модель строится, если необходимо изучить систему согласных какого-либо языка или устройство речевого аппарата человека.
2. ***Функциональные модели*** позволяют изучать поведение некоторого объекта, течение некоторого процесса или же этапы реализации некоторого явления. Например, функциональная модель строится, если необходимо смоделировать процесс создания некоторого текста человеком. Такая же модель создается для объяснения процесса перевода текста с одного языка на другой.
3. ***Динамические модели*** создаются при необходимости найти объяснение некоторых процессов или явлений в их временном развитии. Так, если требуется узнать, как со временем менялось произношение некоторого слова, строят динамическую модель такого процесса.

Особая роль в лингвистике отводится функциональным моделям, позволяющим раскрыть суть функционирования языка, механизма производства и восприятия речи и текста. Нельзя заглянуть в мозг человека и посмотреть, как в нем осуществляются операции с буквами, звуками, словами, предложениями при всевозможном использовании языка. Поэтому для решения таких задач в рамках функциональных моделей выделяют ***воспроизводящие инженерно-лингвистические модели*** (ВИЛМ). Они представляют собой компьютерные системы, поведение которых, с одной стороны, имитирует поведение реальных лингвистических объектов, а с другой стороны, позволяет хотя бы частично воспроизвести эти реальные объекты*.*

Существуют разные способы формализованного описания объекта, процесса или явления: формулы, таблицы, графики, схемы, наборы предложений естественного языка и т.д. Все эти способы составляют основу алгоритмического решения задач с помощью ПК.