

## Общение человека с компьютером (естественный язык и естественность общения)

Наиболее существенными и привлекательными (в контексте задачи общения с компьютером / компьютерной системой) свойствами ЕЯ являются:

- максимально широкое **использование его человеком в своей повседневной деятельности** (это избавляет от необходимости специального изучения формализованного языка общения с ЭВМ и от трудностей, связанных с формулировкой заданий и запросов на таком языке);
- возможность использования естественного языка для **выражения качественно различного содержания** с любой доступной или желательной человеку степенью строгости и полноты (что гарантирует чрезвычайную широту сферы общения – как в плане охвата самых разнообразных предметных областей, так и в плане варьирования формулировок);
- его **открытость** и способность служить метаязыком для самого себя (что обеспечивает расширяемость используемых языковых средств).

Эти обстоятельства (обычно упоминается первое – не только потому, что оно действительно важно, но и потому, что оно абсолютно очевидно, лежит на поверхности) служат очень серьезными доводами в пользу общения с компьютером именно на естественном языке. Пока исследования носили чисто экспериментальный характер, эти доводы были достаточны. Однако в наше время, для которого характерна практическая переориентация работ, возникают новые проблемы, ранее остававшиеся в тени.

Часть из них: необходимость отчуждения системы от разработчика, надежность и устойчивость ее функционирования, эффективность реализации, наличие средств сопровождения – возникает и при создании традиционного программного обеспечения. Новые моменты связаны с использованием для общения с машиной именно естественного языка.

Среди проблем, особо актуальных на нынешнем этапе исследований и разработок, укажем:

- тщательный анализ вопроса целесообразности использования ЕЯ в человеко-машинном общении;
- поиск ситуаций, в которых общение с машиной на ЕЯ оправдано технологически и эргономически;
- выявление обстоятельств, учет которых обеспечивает человеку комфортные, естественные условия общения с компьютером;
- анализ пригодности использовавшихся ранее подходов и методов в изменившихся (практическая переориентация) условиях.

Перед автором некоторого искусственного языка общения с машиной (например, языка программирования), конечно же, не стоит вопрос о целесообразности использования созданного языка по прямому назначению. При оценке такого языка речь может идти о выразительных средствах, эффективной реализуемости, легкости усвоения и т.п. Отдельные неудачные решения могут быть изменены в ходе доработки (и отражены в разного рода пересмотренных сообщениях и др.). Объективация языка заключается в создании стандартов, трансляторов, формировании круга пользователей.

Естественный же язык изначально дан разработчикам компьютерных систем извне, он объективирован (и активно используется в речевой практике) в большой социальной группе носителей данного языка, которые привыкли к вполне определенным, человеческим условиям общения (в том числе, рассмотренным в начале данной главы). Если эти условия (человеческий фактор) будут игнорироваться, язык общения, возможно, сохранив внешнее сходство с тем или иным ЕЯ, потеряет главное – естественность. А учет этих условий требует от разработчиков систем очень серьезных дополнительных усилий, поскольку предполагает воссоздание (моделирование) нетривиальных человеческих механизмов работы с языком, наделение системы – как «собеседника» пользователя – основными чертами (на уровне информационных процессов) собеседника-человека.

Поэтому при создании компьютерных систем практической ориентации следует тщательно проанализировать, оправданы ли интеллектуальные и материальные затраты (весьма значительные, в нынешних условиях отсутствия в нашей стране рынка готового программно-информационного обеспечения) на их разработку, экономична ли (с учетом ресурсоемкости) их эксплуатация.

**Серьезная практическая задача обеспечения общения с ЭВМ на естественном языке требует серьезного и практичного подхода.** В каждой конкретной ситуации необходимо учитывать основательность доводов в пользу общения с системой именно на естественном языке, помнить о реально предоставляемых пользователю удобствах (в частности, об утомительности

клавиатурного ввода, о возможностях – пока весьма скромных – технических средств обеспечения общения: устройства распознавания и синтеза звучащей речи, читающие автоматы).

Стремление разработчика или заказчика не отстать от моды, создать «высокоинтеллектуальную» информационную систему, оснащенную средствами естественноречевого интерфейса, не является достаточно веским основанием, а дилетантский подход (в этой новой и чрезвычайно сложной области особенно) не только не приводит к успеху, но и дискредитирует саму идею общения с ЭВМ на естественном языке.

Рассмотрим особенности естественного языка, осложняющие его использование для общения с ЭВМ (и, разумеется, задачу автоматической обработки текстов):

- ЕЯ – **большая система** (как термин),
- ЕЯ – **иерархическая система**,
- ЕЯ – **открытая система**,
- Связи элементов языка обычно **неоднозначны**,
- Отдельные носители языка используют **индивидуальные модели языка**,
- Использование языка сопровождается **речевыми ошибками**,
- **Описания языка** (построенные специалистами – лингвистами) **не полны и не точны**.

Задачу обеспечения **естественного общения** человека с машиной можно принять без каких бы то ни было оговорок. Однако ниоткуда не следует, что наиболее удобным и естественным для пользователя (и целесообразным, с точки зрения разработчика) средством такого общения будет естественный язык. Пререкания с «непонятливым» компьютером, ориентированным на ведение диалога с человеком на так называемом ограниченном естественном языке, могут потребовать более значительных усилий, чем изучение искусственного формального языка общения.

Нас интересуют ситуации, в которых необходимость использования естественного языка диктуется глубинными внутренними причинами:

- характер поручаемых системе заданий, а следовательно, и адресуемых ей сообщений таков (носит предварительный, неформальный характер), что описать их на каком-либо формализованном языке крайне трудно;

- общение пользователя с машиной происходит эпизодически и/или в очень широкой сфере (изучение специального языка общения нецелесообразно, или же он становится необъятным).

Часто в таких случаях альтернативой использованию естественного языка может служить хорошо разработанная схема «выспрашивания» нужной системе информации с помощью традиционных средств организации дружественного интерфейса (меню, опережающий ввод и др.).

Весьма интересен феномен появления **гибридных знаковых систем**.

Мы знаем, что конструкции естественного языка (слова, словосочетания, предложения, тексты) давно и широко используются в человеко-машинном интерфейсе.

При формулировке общих требований к диалогу часто отмечается, что диалог должен вестись на родном языке пользователя (или на другом понятном и привычном ему языке).

Вспомним:

- служебные слова в языках программирования, командных и других языках;
- названия позиций меню, кнопок и т.п.;
- средства поддержки пользователя:
  - сообщения об ошибках,
  - справочная информация,
  - внешняя документация.

Можно ли в подобных ситуациях говорить об общении с ЭВМ на естественном языке?

Вряд ли. Средства общения с традиционными компьютерными системами требуют предварительной экспликации, формализации той ситуации, о которой идет речь, привлечения знаний об используемых в системе способах структуризации знаний. Формулируя же адресуемые машине сообщения на естественном языке, человек может не знать: каким набором семантических единиц располагает компьютерная система, в какие формальные структуры будет отображаться содержание сообщения; какие компоненты описываемой ситуации являются в настоящий момент значимыми, а какие – второстепенными.

Более того, оказавшись в абсолютно новых условиях, человек, как правило, не сможет воспользоваться штатными языками общения с компьютером. Эти обстоятельства могут оказаться решающими при выборе языка общения с машиной как для профессионала (который на очередном этапе решения задачи не располагает пока адекватной формальной моделью и вынужден по этой

причине отказаться от привычного для него искусственного языка), так и для конечного пользователя (для которого содержательный уровень общения является единственно доступным).

#### **Лингвистическое и алгоритмическое обеспечение общения с компьютером на естественном языке.**

Для того, чтобы привлекательность диалога с компьютером на естественном языке не просто декларировалась, а стала реальностью, необходимо не только тщательно проанализировать условия и сценарии общения, но и:

- построить формальную модель естественного языка (учитывающую все его ключевые особенности),
- описать нетривиальное подмножество естественного в рамках этой модели,
- разработать и реализовать алгоритмы анализа и синтеза текстов.

Один из главных путей развития функциональных возможностей систем общения и повышения качества их работы – создание и внедрение более полных и точных моделей естественных языков, более совершенных алгоритмов анализа и синтеза текста.

#### **Лингвистические банки данных.**

Под *лингвистическими банками данных* (ЛБД) понимаются представленные в электронной форме языковые источники (корпусы текстов) и лингвистические описания. Отметим, что в наше время, в ситуации, когда надежность работы систем оптического распознавания близка (на хороших по качеству печатных текстах) к 100%, в электронную форму легко переводимы и традиционные источники информации о языке. Поэтому можно считать, что в ЛБД можно перевести любые полиграфические источники: тексты на том или ином естественном языке, словари, справочники, книги по лингвистике. Спектр ЛБД достаточно широк: это как необработанные («сырые») корпусы текстов, так и тексты с некоторыми добавлениями, например грамматическими характеристиками слов, стилистическими пометами (разговорное, специальное и т.п.), или описаниями синтаксической структуры предложений (соответствующие корпусы текстов называют *размеченными*). Сюда также входят разнообразные компьютерные словари: частотные, грамматические, словоформ, тезаурусы, словари словосочетаний и моделей управления, своды грамматических правил и т.п.

Различаться может и назначение лингвистических банков данных. Часть ЛБД предназначена для автоматизации деятельности лингвистов и разработчиков прикладных систем, часть - для непосредственного использования в системах обработки текста и речи: автокорректоры, системах распознавания текста и речи, информационно-поисковых системах.

Два типа технологий формирования ЛБД:

- Формирование "вручную" экспертами (филологи, специалисты в области применения компьютерной системы); здесь нужны инструментальные средства поддержки их работы.
- Формирование в автоматизированном режиме с использованием методов машинного обучения; желательно участие экспертов для анализа и коррекции результатов, полученных в автоматическом режиме.