Natural Language Processing

INTRODUCTION

Husni Al-Muhtaseb

Tuesday, February 20, 2007

بسم الله الرحمن الرحيم
ICS 482: Natural Language Processing

INTRODUCTION

Husni Al-Muhtaseb

Tuesday, February 20, 2007

Course Description

- To introduce students to different issues concerning the creation of computer programs that can interpret, generate, and learn natural language. Among the issues that will be discussed are: syntactic processing, semantic interpretation, discourse processing, knowledge representation and the acquisition of grammatical and lexical knowledge. The primary emphasis of this course is on text-based language processing (not speech).

Prerequisite

- Senior Standing in ICS major

- Mastering at least one programming language

Instructor

- ***Name*: *Husni Al-Muhtaseb***

- ***Office*: *Bldg. 22 Room 311***
  Phone #: *2624 *Email*:
  muhtaseb@kfupm.edu.sa

- ***web*: http://faculty.kfupm.edu.sa/ics/muhtaseb/**

Office Hours

- *Sunday, Monday & Tuesday* 11:20 -11:50
- *Sunday, Monday & Tuesday* 12:20 – 01:00

Electronic mail

- External one
- Clear name. Sign always
  - blackbird@fly.net or Ali_Sami@fly.net
- Shouting: SALAM virsus salam
- Symbols
  - :-)   ☺
  - :-(☹
- Read before sending

Husni Al-Muhtaseb

Online course site

- *http://webcourses.kfupm.edu.sa/*
- Material & Notes
- Assignments and Submission (Assign. 1 is there)
- Discussions & participation
- Mail
- Grades

Grading Policy

Category	Weight
Assignments	0%
Quizzes (4)	28%
Project	25%
Presenting a Topic	10%
Participation	12%
Final Exam	25%
Total	100 %

Quizzes

- 30 minutes

- Announced at least 2 days before

- In class time

Textbook

- *Speech And Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, By Daniel Jurafsky and James H. Martin, Prentice-Hall, 2000. http://www.cs.colorado.edu/~martin/slp.html

- Several Chapters have been re-written & re-numbered

  - Visit Book website

Tentative Weekly Schedule

W#	Topic	Textbook Chapters	Activity
1	Introduction	1	
2	Regular Expressions & Automata	2	
3	Morphology & Finite State Transducers	3	
4	N-Grams	6	Quiz 1
5	Parts of Speech	8 + external Material	
6	Syntax & Context-free grammars - Parsing	9 & 10	
7	Lexicalized and Probabilistic Parsing	11	Quiz 2

Tentative Weekly Schedule

W#	Topic	Chapters	Activity
8	Semantic Representation & Representing Meaning	14	
9	Semantic analysis & lexical Semantics	15 & 16	
10	Wrap up		Quiz 3
11	Machine Translation	21	
12	Information Extraction	Ext. Mat.	
13-15	Students' presentations		Quiz 4 - Take-home

Questions

- NLP: Natural Language Processing
- NLU: Natural Language Understanding
- NLC: Natural Language Computing
- HLP: Human Language Processing
- HLU: Human Language Understanding
- HLC: Human Language Computing
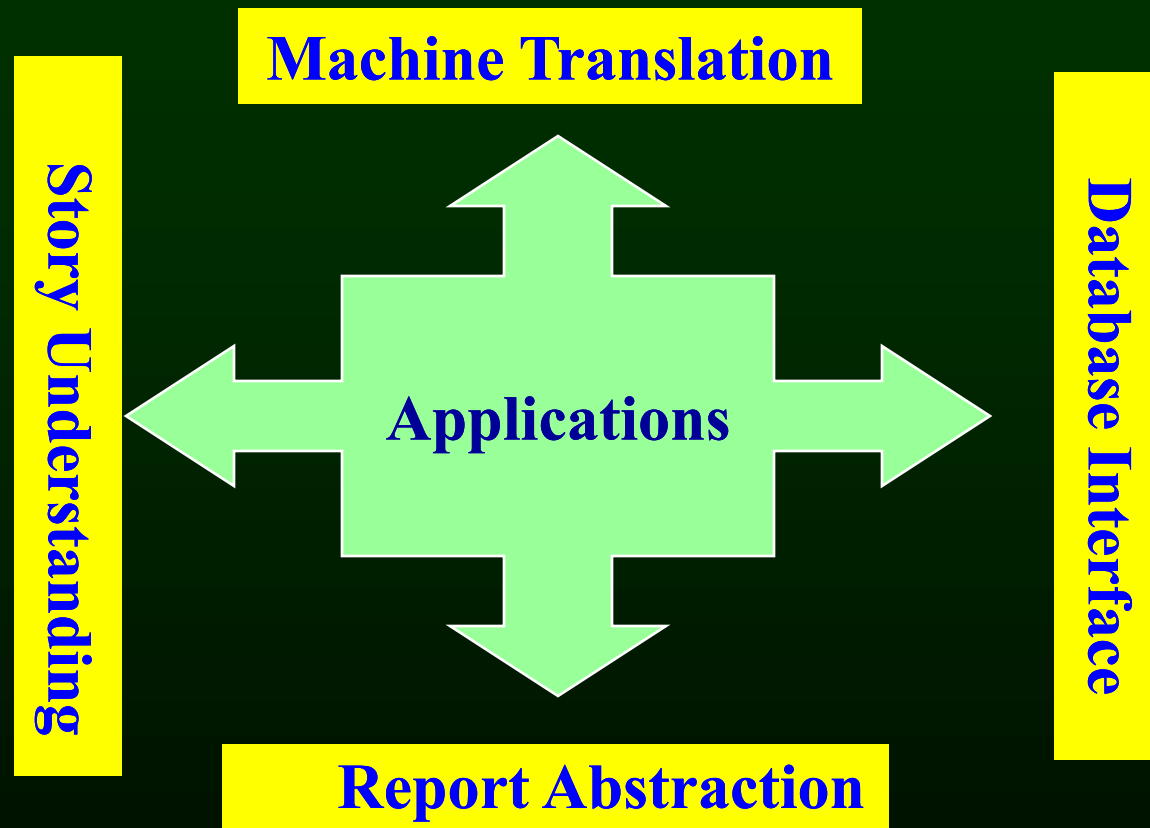- CL: Computational Linguistics

Definitions

NLP

The sub-domain of artificial intelligence concerned with the task of developing programs possessing some capability of '**understanding**' a natural language in order to achieve some specific goal

A transformation from one representation (*the input text*) to another (*internal representation*)
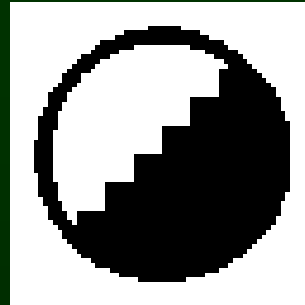
understanding

Motivation

Machine Translation

Story Understanding

Applications

Database Interface

Report Abstraction

Husni Al-Muhtaseb

Stages of NLP

Morphological Analysis
Individual words are analyzed into their components

Syntactic Analysis
Linear sequences of words are transformed into structures that show how the words relate to each other

Semantic Analysis
A transformation is made from the input text to an internal representation that reflects the meaning

Pragmatic Analysis
To reinterpret what was said to what was actually meant

Discourse Analysis
Resolving references Between sentences

Husni Al-Muhtaseb

The Steps in NLP

Discourse

Pragmatics

Semantics

Syntax

Morphology

**we can go up, down and up and down and combine steps too!!

**every step is equally complex

Husni Al-Muhtaseb

The steps in NLP (Cont.)

- <u>Morphology</u>: Concerns the way words are built up from smaller meaning bearing units.

- <u>Syntax</u>: concerns how words are put together to form correct sentences and what structural role each word has

- <u>Semantics</u>: concerns what words mean and how these meanings combine in sentences to form sentence meanings

The steps in NLP (Cont.)

- <u>Pragmatics</u>: concerns how sentences are used in different situations and how use affects the interpretation of the sentence

- <u>Discourse</u>: concerns how the immediately preceding sentences affect the interpretation of the next sentence

Parsing (Syntactic Analysis)

- Assigning a syntactic and logical form to an input sentence

  - uses knowledge about word and word meanings (lexicon)
  - uses a set of rules defining legal structures (grammar)

- Ahmad ate the apple.

```
(S  (NP (NAME Ahmad))
    (VP  (V ate)
        (NP  (ART the)
            (N apple))))
```

Word Sense Resolution

- Many words have many meanings or senses

- We need to resolve which of the senses of an ambiguous word is invoked in a particular use of the word

- I made her duck. (made her a bird for lunch or made her move her head quickly downwards?)

Reference Resolution

- Domain Knowledge (Registration transaction)
- Discourse Knowledge
- World Knowledge
  - U: I would like to register in an IAS Course.
  - S: Which number?
  - U: Make it 333.
  - S: Which section?
  - U: Which section starts at 7:00 am?
  - S: section 5.
  - U: Then make it that section.

Example

Surface form

I want to print Ali's .init file

Morphological analysis

stems

I (pronoun)
want (verb)
to (prep)
to (infinitive)
print (verb)
Ali (noun)
's (possessive)
.init (adj)
file (noun)
file (verb)

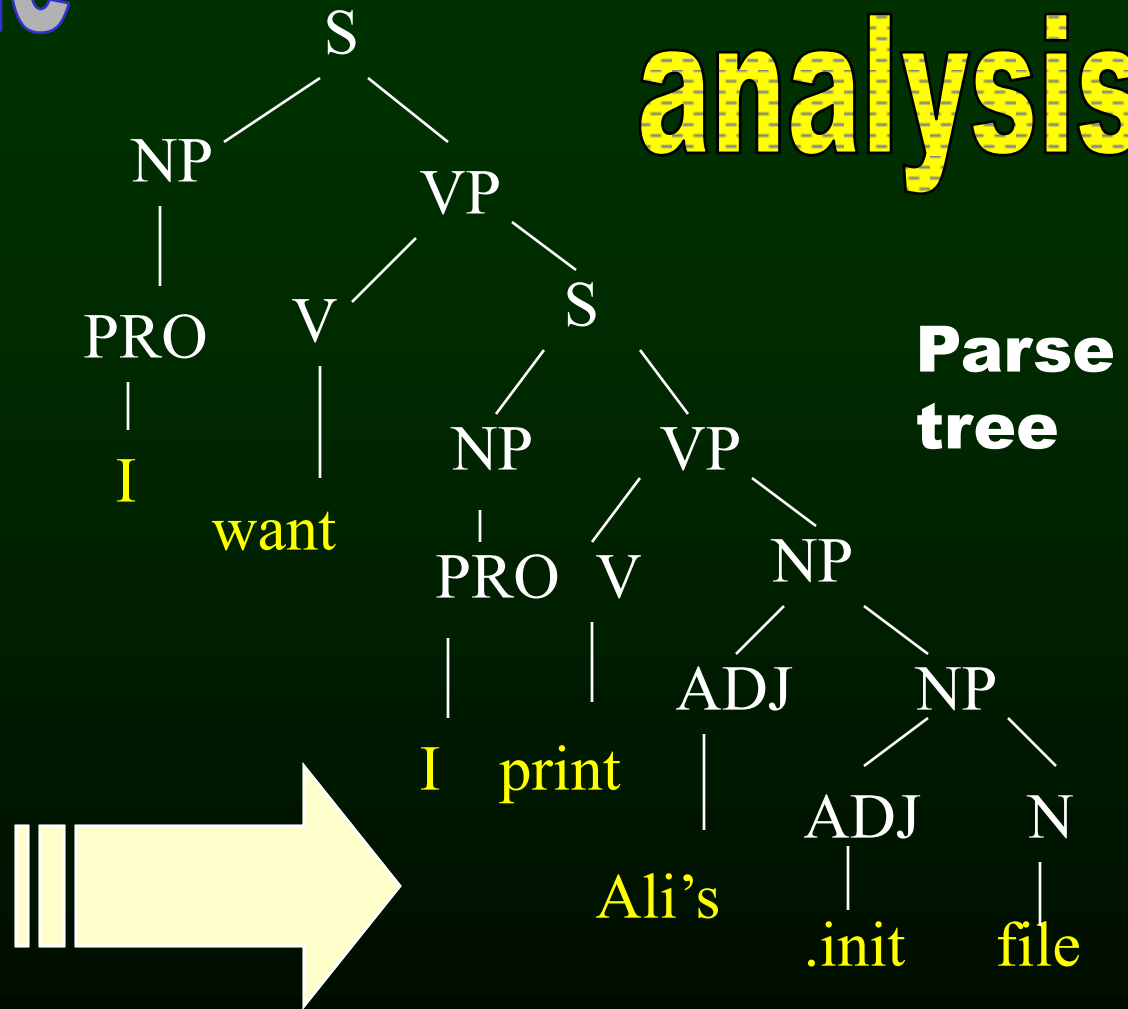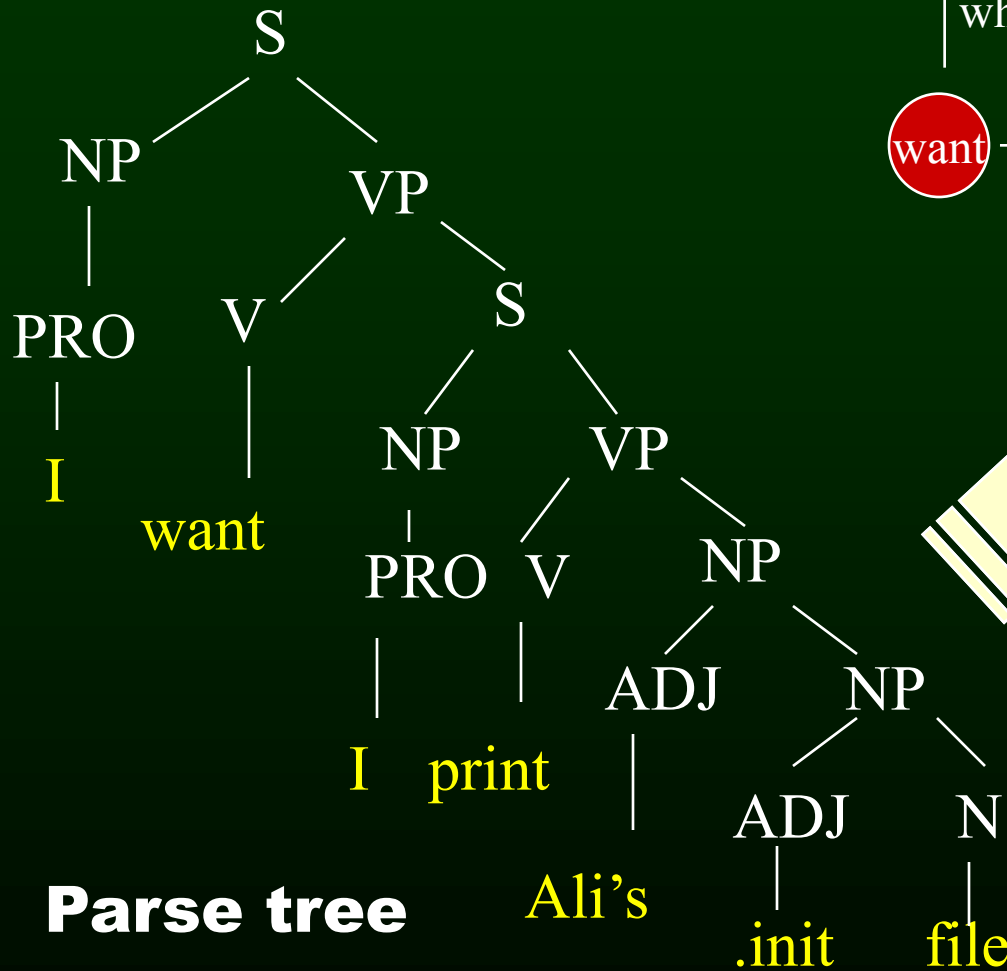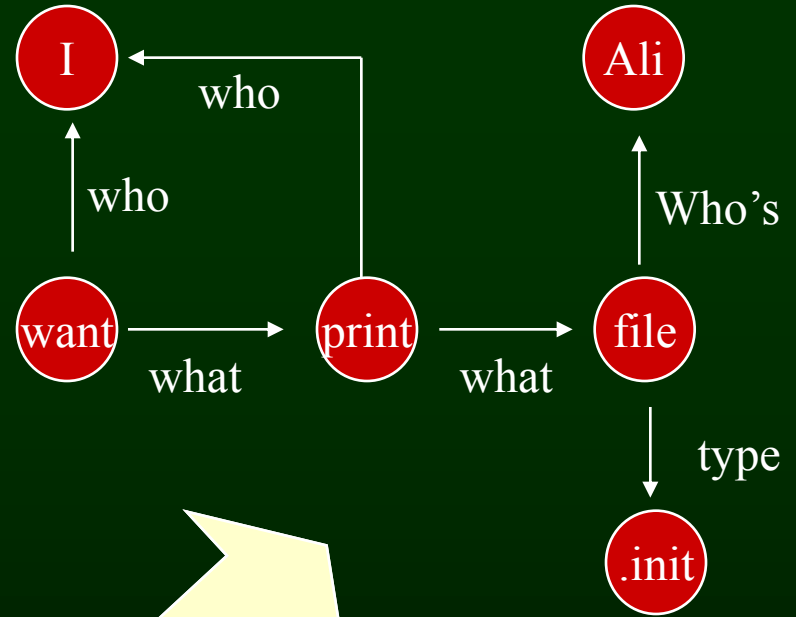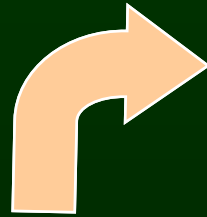# Example

Syntactic analysis

stems

- I (pronoun)
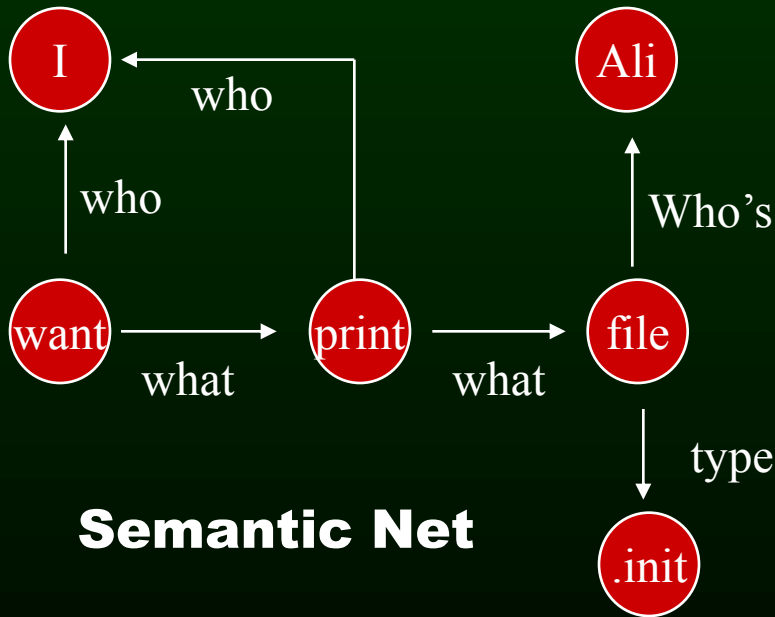- want (verb)
- to (prep)
- to(infinitive)
- print (verb)
- Ali (noun)
- 's (possessive)
- .init (adj)
- file (noun)
- file (verb)

Parse tree

```
                    S
           _____/ _____
          NP                  VP
          |              _____/ \_____
          PRO           V             S
          |             |         ____/ \____
          I           want       NP          VP
                                 |       ____/ \____
                                PRO     V           NP
                                 |      |        ___/ \___
                                 I    print    ADJ        NP
                                              |          /  \
                                            Ali's      ADJ    N
                                                       |      |
                                                     .init   file
```

Example

Semantic Net

Parse tree

```
                S
              /   \
            NP      VP
            |      /  \
           PRO    V     S
            |     |    /  \
            I   want  NP    VP
                      |    /  \
                     PRO  V     NP
                      |   |    /   \
                      I print ADJ    NP
                              |     /  \
                            Ali's  ADJ   N
                                    |    |
                                  .init file
```

Semantic Net

- I
- who
- who
- Ali
- Who's
- want → (what) → print → (what) → file
- file → (type) → .init

Semantic analysis

Example

Discourse →

To whom the pronoun 'I' refers

To whom the proper noun 'Ali' refers
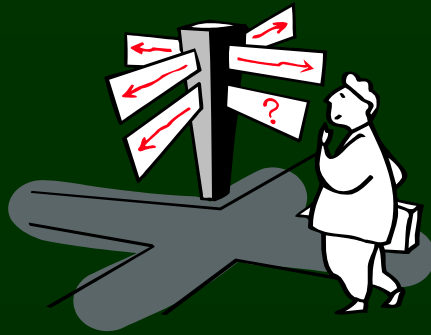
What are the files to be printed

Semantic Net

- I
- Ali

who

who

who

Who's

want → print → file

what   what

type

.init

Pragmatic

Execute the command

`lpr /ali/stuff.init`

Stages of NLP

Syntactic Analysis

stems

Morphological Analysis

parse tree

user

Semantic Analysis

lexicon

Surface form

Internal representation

Discourse Analysis

Resolve references

Pragmatic Analysis

Perform action
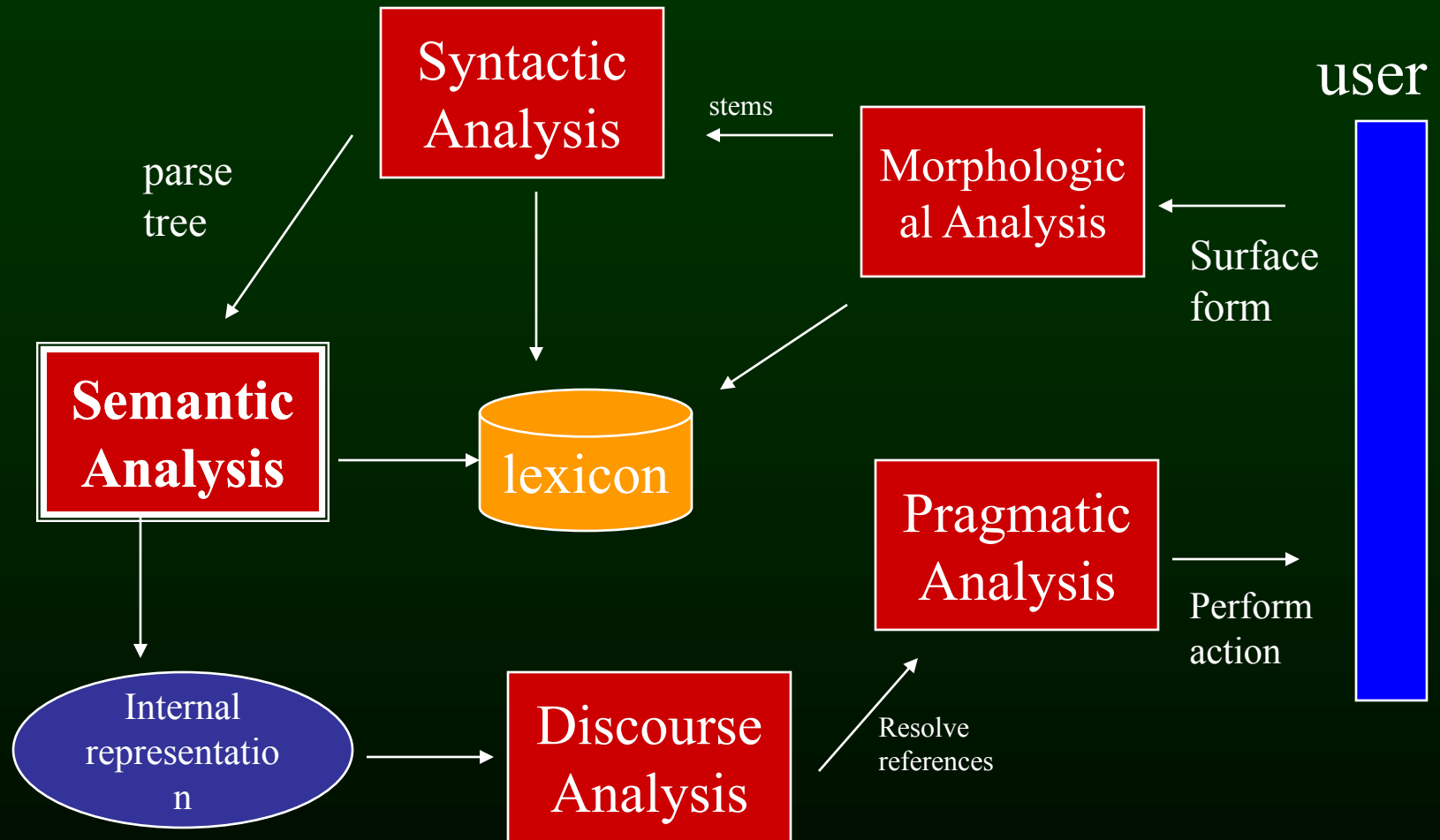
Ambiguity

more than one meaning for the same sentence

Time flies like an arrow

Time passes along in the same manner as an arrow gliding through space.

I order you to take timing measurements on flies, in the same manner as you would time an arrow. (other different meanings)
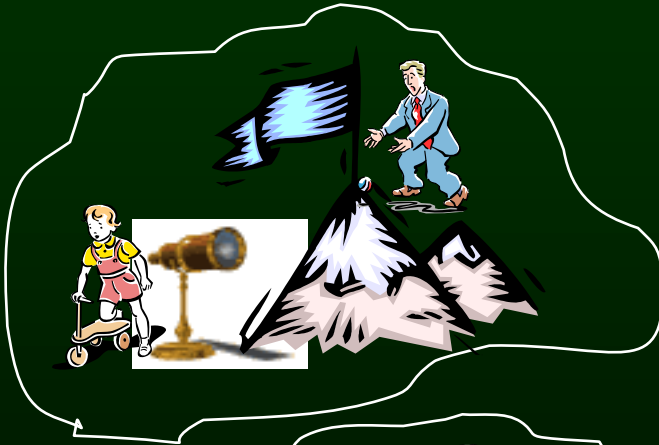
Fruit flies like to feast on a banana; in contrast, the species of flies known as "time flies" like an arrow.

Ambiguity

The boy saw the man on the mountain with a telescope

Prepositional phrase attachment

Husni Al-Muhtaseb
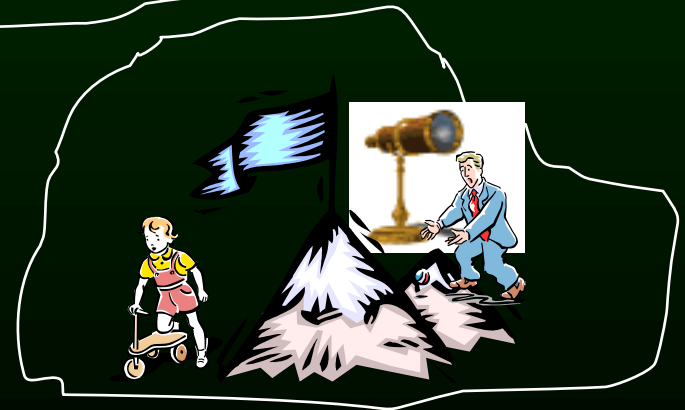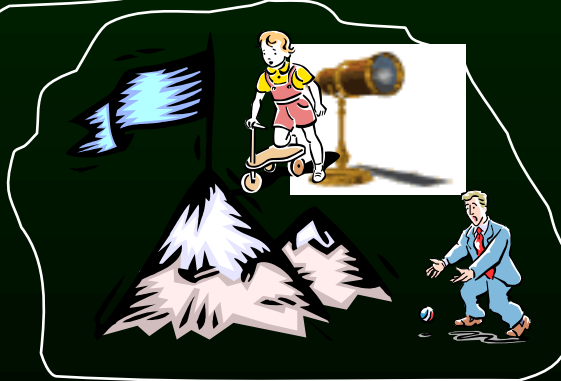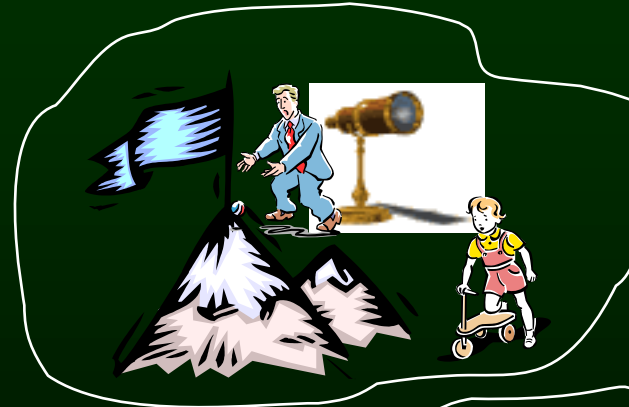
Short men and women

Visiting relatives can be boring

I made her duck

Bengali history teacher

We eat what we can, and what we can not we can
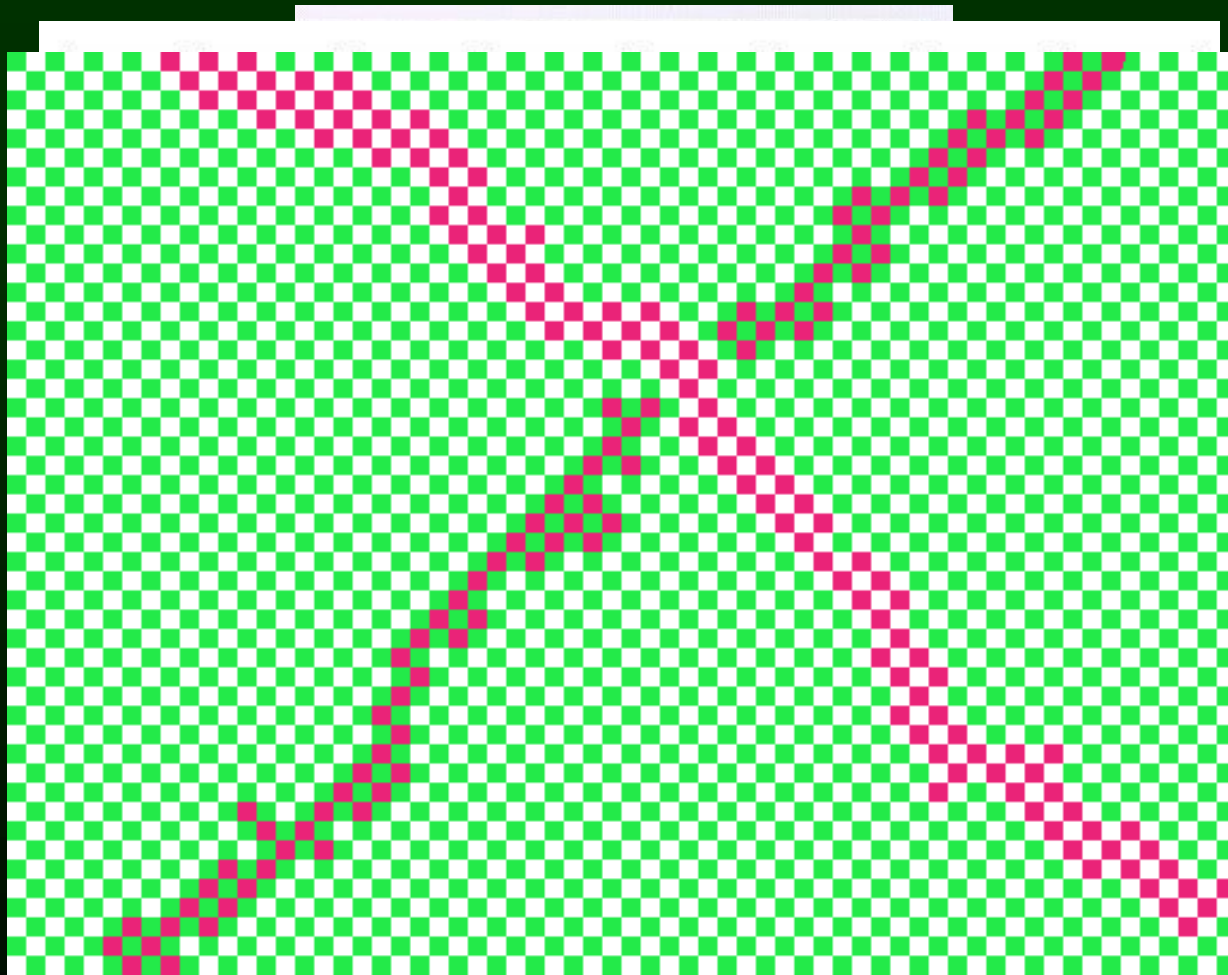
Ambiguity

Ali knows a richer man than Ahmad

Record the signal strength at the wire under normal load

The chicken is ready to eat

Ambiguity

Ambiguity

A program

Lexicon

- Lexicon is a vocabulary data bank, that contains the language words and their linguistic information.

- •There are many on-line lexicon

  WordNet is a lexical database that contains English vocabulary words

- **COULD WE HAVE ONE FOR ARABIC?**

Simple Applications

- Word counters (wc in UNIX)
- Spell Checkers, grammar checkers
- Predictive Text on mobile handsets

Bigger Applications

- Intelligent computer systems
- NLU interfaces to databases
- Computer aided instruction
- Information retrieval
- Intelligent Web searching
- Data mining
- Machine translation
- Speech recognition
- Natural language generation
- Question answering

Spoken Dialogue System
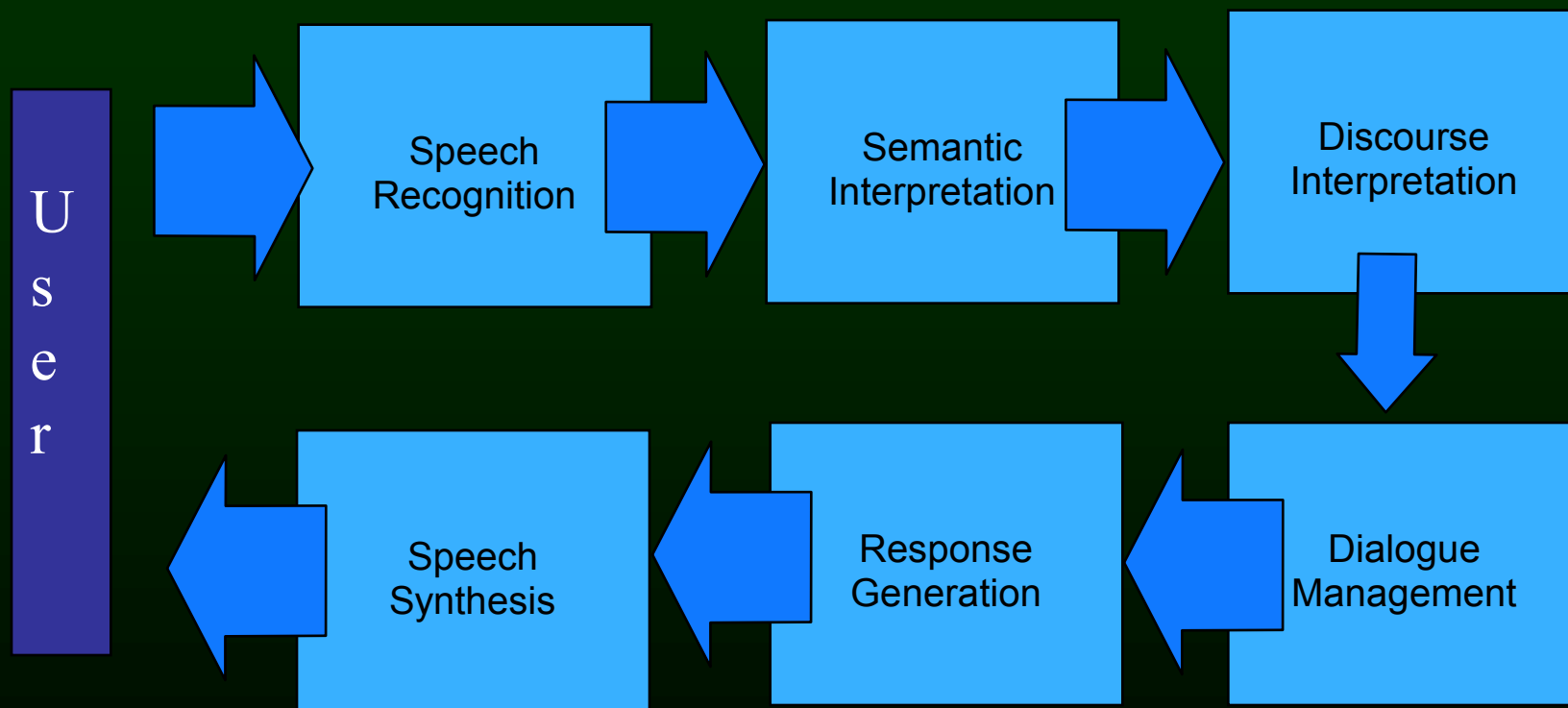
Parts of the Spoken Dialogue System

- Signal Processing: Convert the audio wave into a sequence of feature vectors.
- Speech Recognition: Decode the sequence of feature vectors into a sequence of words.
- Semantic Interpretation: Determine the meaning of the words.
- Discourse Interpretation: Understand what the user intends by interpreting utterances in context.
- Dialogue Management: Determine system goals in response to user utterances based on user intention.
- Speech Synthesis: Generate synthetic speech as a response.

Levels of Sophistication in a Dialogue System

- Touch-tone replacement:

  System Prompt: "For checking information, press or say one."
  Caller Response: "One."

- Directed dialogue:

  System Prompt: "Would you like checking account information or rate information?"
  Caller Response: "Checking", or "checking account," or "rates."

- Natural language:

  System Prompt: "What transaction would you like to perform?"
  Caller Response: "Transfer Rs. 500 from checking to savings."

Thank you