CSC 480: Artificial Intelligence

Franz J. Kurfess

Visiting Professor
Department of Computer Science and Mathematics
Munich University of Applied Sciences
Germany

Professor
Computer Science Department
California Polytechnic State University
San Luis Obispo, CA, U.S.A.

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Course Overview

- **Introduction**

- **Intelligent Agents**

- **Search**
  - problem solving through search
  - uninformed search
  - informed search

- **Games**
  - games as search problems

- **Knowledge and Reasoning**
  - reasoning agents
  - propositional logic
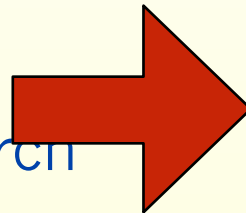 - predicate logic
  - knowledge-based systems

- **Learning**

- PAC learning
- learning from observation
- neural networks

- **Natural Language Processing (NLP)**

- **Conclusions**

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Chapter Overview
NLP

- ❖ **Motivation**

- ❖ **Objectives**

- ❖ **Introduction NLP**
  - ❖ Terminology
  - ❖ Related Areas

- ❖ **Communication**
  - ❖ Roles: Speaker, Hearer
  - ❖ Communication Channels

- ❖ **NLP Perspectives**
  - ❖ Syntax
  - ❖ Semantics
  - ❖ Pragmatics

- ❖ **Core NLP Issues**
  - ❖ Modularity

- ❖ Ambiguity

- ❖ **NLP Tasks**
  - ❖ Syntactic Tasks
  - ❖ Semantic Tasks
  - ❖ Pragmatic Tasks

- ❖ **Related Tasks**
  - ❖ Information Extraction
  - ❖ Text Summarization
  - ❖ Machine Translation
  - ❖ Knowledge Acquisition

- ❖ **History of NLP**

- ❖ **Important Concepts and Terms**

- ❖ **Chapter Summary**

Introduction NLP

Terminology
Related Areas

Natural vs. Artificial Language

❖ **natural languages**
  ❖ primarily used by people
    ❖ spoken language
    ❖ written language
  ❖ evolved over time
  ❖ frequently with complex, but flexible grammatical structures

❖ **artificial languages**
  ❖ designed and created for specific uses
    ❖ "neutral" language independent of nationalistic or ethnic influences
      ❖ Esperanto
      ❖ Lojban
  ❖ some are primarily intended for computers
    ❖ programming languages
    ❖ knowledge representation languages
  ❖ grammatical structures are often simple, but rigid

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN·FH
MÜNCHEN

Related Areas

❖ **Computational Linguistics**
  ❖ use of computational methods in linguistics

❖ **Information and Knowledge Extraction**
  ❖ analysis of texts

❖ **Formal Languages**
  ❖ artificial languages
    ❖ emphasis on certain abstract properties

❖ **Cognitive Science**
  ❖ language and its use by humans

❖ **Machine Learning**
  ❖ acquisition of language properties from data sets

Communication

**Roles: Speaker, Hearer
Communication Channels**

Communication

❖ **exchange of information between two or more participants**
- ❖ requires a common communication infrastructure and method

❖ **communication channel(s)**
- ❖ infrastructure for the information exchange
- ❖ physical communication medium
  - ❖ sound over the air for spoken language

❖ **alphabet**
- ❖ set of symbols comprehensible to speaker and hearer

❖ **language**
- ❖ method for constructing longer sequences of symbols for richer and more effective communication

❖ **discourse**
- ❖ conventions and rules about behavior during communication

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Communication: Speaker

❖ **Intention**

❖ **Generation**

❖ **Synthesis**

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Communication: Hearer

- ❖ **Perception**

- ❖ **Analysis**

- ❖ **Incorporation**

Transfer of Knowledge

Communication

Basic Concepts

Language and Communication

Natural Language

Formal Languages

Communication Models

Basic Concepts

- **communication**
  - exchange of information
  - requires a shared system of signs
  - greatly enhanced by language
  - speaker
    - produces signs as utterances
    - general: not only spoken language
  - listener (hearer)
    - perceives and interprets signs

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Purpose of Communication

❖ **sharing of information among agents or systems**
  ❖ query other agents for information
  ❖ responses to queries
  ❖ requests or commands
    ❖ actions to be performed for another agent
  ❖ offer
    ❖ proposition for collaboration
  ❖ acknowledgement
    ❖ confirmation of requests, offers
  ❖ sharing
    ❖ of experiences, feelings

Communication Problems

❖ **intention**
  ❖ what is the expected outcome (speaker's perspective)

❖ **timing**
  ❖ when is a communication act appropriate

❖ **selection**
  ❖ which act is the right one

❖ **language**
  ❖ what sign system should be used

❖ **interpretation**
  ❖ will the intended meaning be conveyed to the listener

❖ **ambiguity**
  ❖ can the intention be expressed without the possibility of misunderstandings

© Franz J. Kurfess

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Language and Communication

❖ **Natural Language**
  ❖ used by humans
  ❖ evolves over time
  ❖ moderately to highly ambiguous

❖ **Formal Languages**
  ❖ invented
  ❖ rigidly defined
  ❖ little ambiguity

Natural Language

❖ **formal description is very difficult**
  ❖ sometimes non-systematic, inconsistent, ambiguous

❖ **mostly used for human communication**
  ❖ easy on humans
  ❖ tough on computers

❖ **context is critical**
  ❖ situation, beliefs, goals

16

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

CAL POLY

Formal Languages

- **symbols**
  - terminal symbols
    - finite set of basic words
    - not: alphabet, characters
  - non-terminal symbols
    - intermediate structures composed of terminal or non-terminal symbols

- **strings**
  - sequences of symbols

- **phrases**
  - sub-strings grouping important parts of a string

Formal Languages Cont.

❖ **sentences**
  ❖ allowable strings in a language
  ❖ composed from phrases

❖ **grammar**
  ❖ rules describing correct sentences
  ❖ often captured as rewrite rules in BNF notation

❖ **lexicon**
  ❖ list of allowable vocabulary words

Communication Models

❖ **encoded message model**
  ❖ a definite proposition of the speaker is encoded into signs which are transmitted to the listener
  ❖ the listener tries to decode the signs to retrieve the original proposition
  ❖ errors are consequences of transmission problems

❖ **situated language model**
  ❖ the intended meaning of a message depends on the signals as well as the situation in which they are exchanged
  ❖ mis-interpretation may lead to additional problems
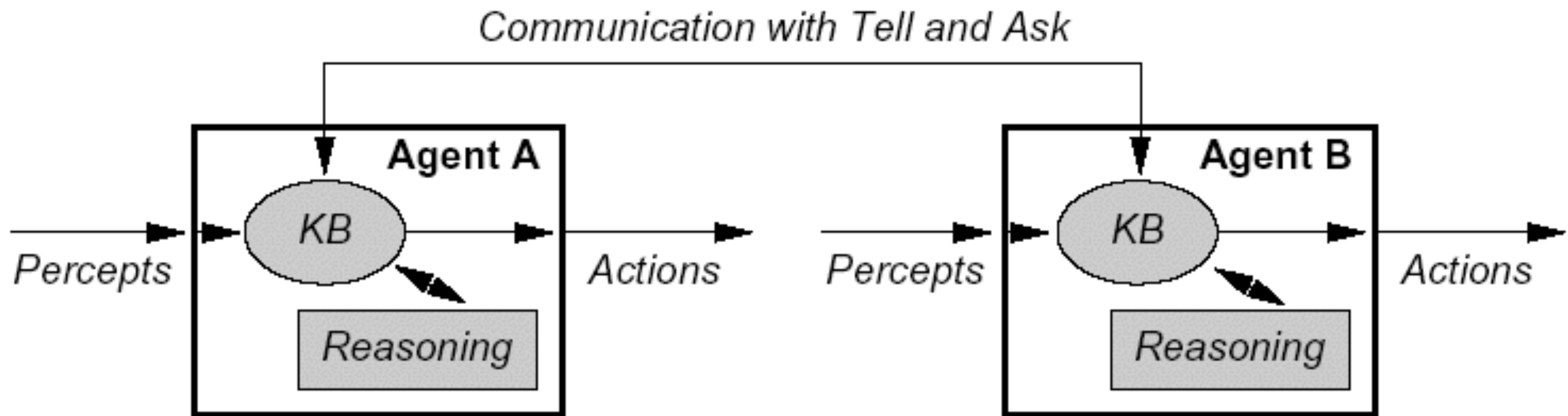
Communication Types

❖ **telepathic communication**
  ❖ speaker and listener have a shared internal representation
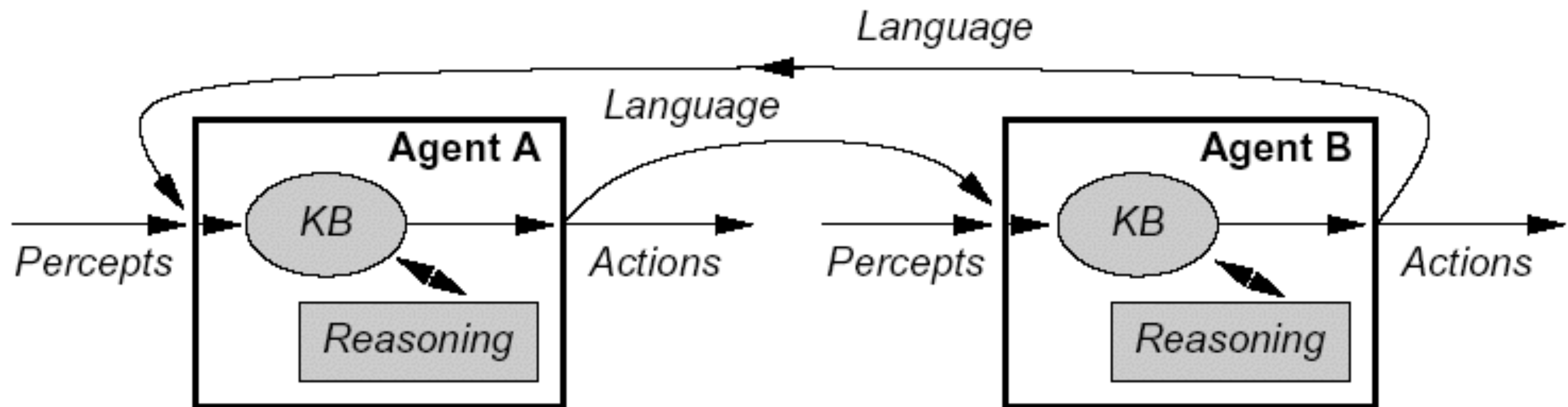 ❖ communication through Tell/Ask directives

❖ **language-based communication**
  ❖ speaker performs actions that produce signs which other agents can perceive and interpret
  ❖ communication language is different from the internal representation
  ❖ more complex
    ❖ involves several mappings
    ❖ language needs to be generated, encoded, transmitted, decoded, and interpreted

Telepathic Communication

Communication with Tell and Ask

Agent A — KB — Reasoning
Percepts — Actions

Agent B — KB — Reasoning
Percepts — Actions

[Russell & Norvig 1995]

Language-Based Communication

[Russell & Norvig 1995]

Communication Steps: Speaker

- ❖ **intention**
  - ❖ decision about producing a speech act

- ❖ **generation**
  - ❖ conversion of the information to be transferred into the chosen language

- ❖ **synthesis**
  - ❖ actions that produce the generated signs

Communication Steps: Hearer

❖ **perception**
  ❖ reception of the signs produced by the speaker
    ❖ speech recognition, lip reading, character recognition

❖ **analysis**
    ❖ syntactic interpretation (parsing)
    ❖ semantic interpretation
  ❖ disambiguation
    ❖ selection of the most probable intended meaning

❖ **incorporation**
    ❖ the selected interpretation is added to the existing world model as additional piece of evidence

24

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Communication Example

[Russell & Norvig 1995]

NLP Perspectives

NLP Levels
Syntax
Semantics
Pragmatics
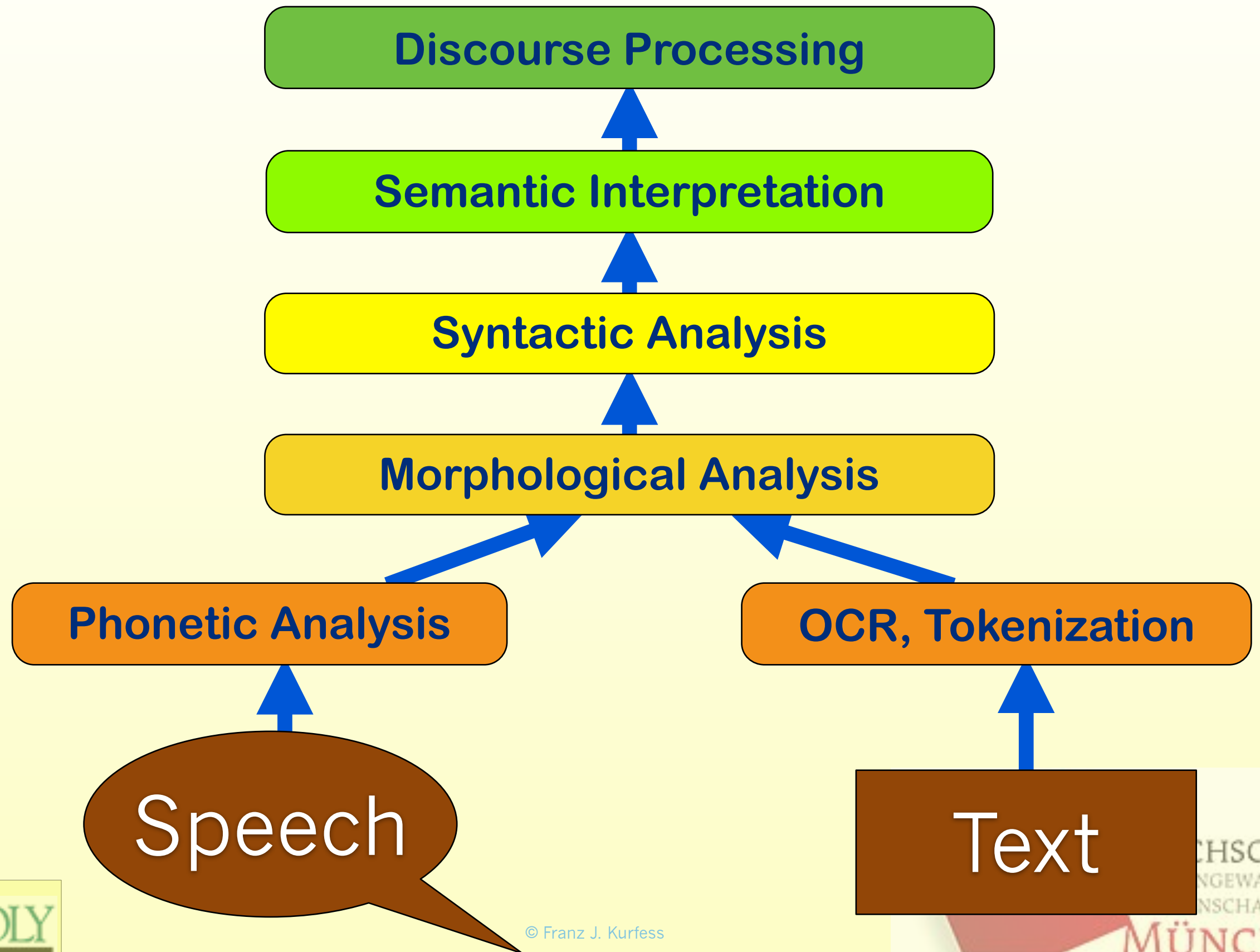
Syntax

❖ **arrangement of words into longer structures**
  - ❖ phrases
  - ❖ sentences

❖ **specified by a grammar**
  - ❖ set of rules about admissible structures
  - ❖ see also syntax check in programming languages
  - ❖ natural languages are difficult to fully describe through a grammar
    - ❖ in particular spoken languages

❖ **helps resolve the meaning of words**
  - ❖ place or role in a sentence

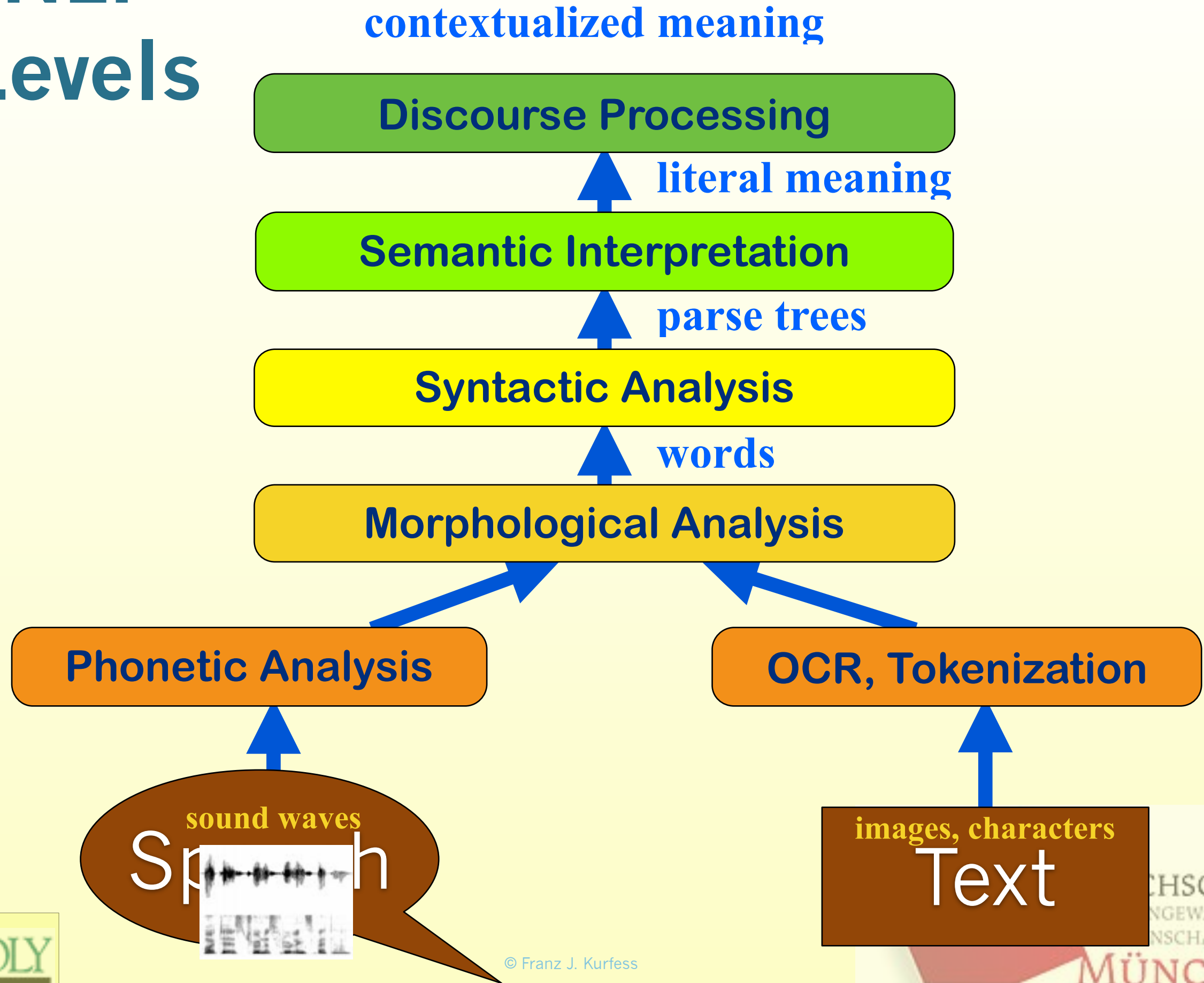# Semantics

❖ **meaning of structures**
  ❖ words
  ❖ phrases
  ❖ sentences

❖ **sometimes specified by (formal) interpretations**
  ❖ mappings of words and phrases to meanings
  ❖ may not be practical for natural languages

❖ **difficult to capture for computers**

Pragmatics

❖ **context and social structures of a communication**
  ❖ provides additional help with the interpretation

Core NLP Issues

Modularity
Ambiguity

Modularity

❖ **compositionality is desirable**
  - ❖ isolated processing of smaller building blocks
  - ❖ assembly of larger blocks from smaller ones
  - ❖ possible for some (artificial) languages
    - ❖ context-free languages
  - ❖ very limited in natural language
    - ❖ context is required for interpretation of words, phrases, sentences, paragraphs, ...
    - ❖ sometimes even for morphological analysis

Ambiguity

❖ **multiple meanings for one structure**
  ❖ word, phrase, sentence, paragraph, ...

❖ **inherent in all natural languages**
  ❖ often used for creative purposes
  ❖ e.g. poetry, puns

❖ **disambiguation is needed**
  ❖ determines the intended meaning
  ❖ usually dependent on the context
    ❖ semantic
    ❖ pragmatic
  ❖ may require "world knowledge" or "commonsense reasoning"

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN·FH
MÜNCHEN

Ambiguity Examples

- Natural language is highly ambiguous and must be *disambiguated*.
  - I saw the man on the hill with a telescope.
  - I saw the Grand Canyon flying to LA.
  - Time flies like an arrow.
  - Horse flies like a sugar cube.
  - Time runners like a coach.
  - Time cars like a Porsche.
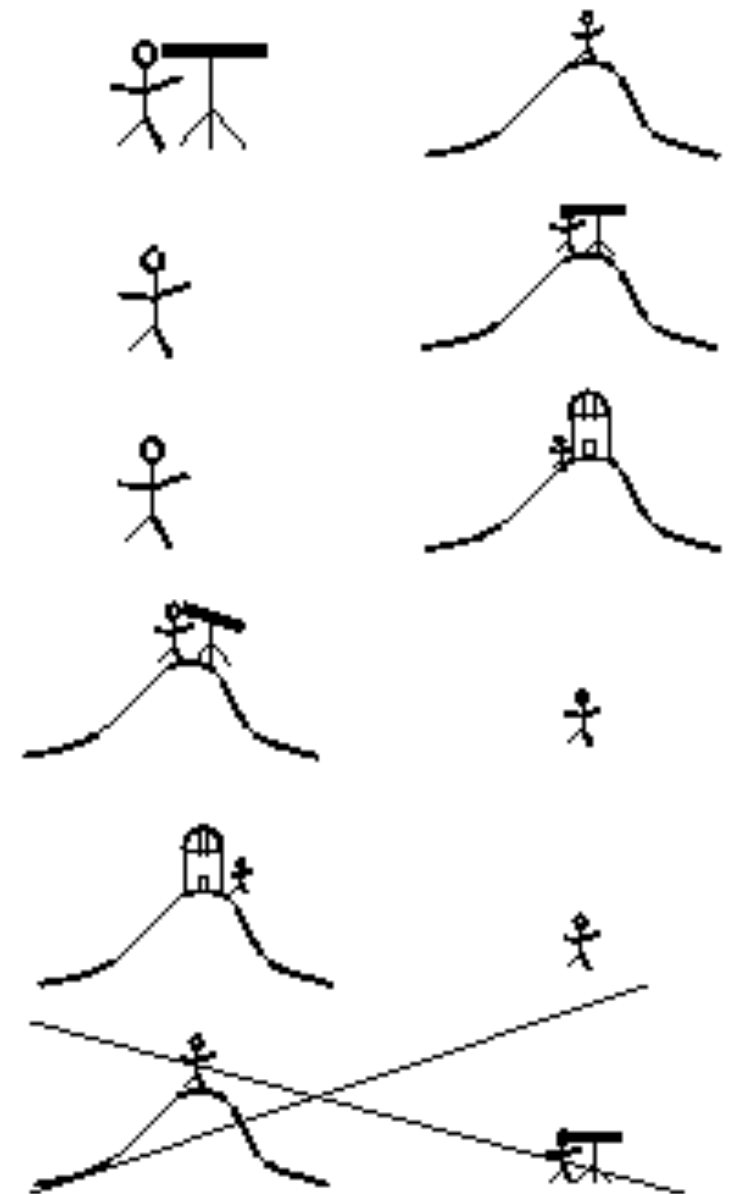
Ambiguity is Ubiquitous

- ## Speech Recognition
  - "recognize speech" vs. "wreck a nice beach"
  - "youth in Asia" vs. "euthanasia"
- ## Syntactic Analysis
  - "I ate spaghetti with chopsticks" vs. "I ate spaghetti with meatballs."
- ## Semantic Analysis
  - "The dog is in the pen." vs. "The ink is in the pen."
  - "I put the plant in the window" vs. "Ford put the plant in Mexico"
- ## Pragmatic Analysis
  - **From "The Pink Panther Strikes Again":**
  - **Clouseau**: Does your dog bite?
    Hotel Clerk: No.
    Clouseau: [*bowing down to pet the dog*] Nice doggie.
    [*Dog barks and bites Clouseau in the hand*]
    Clouseau: I thought you said your dog did not bite!
    Hotel Clerk: That is not my dog.

Ambiguity is Explosive

- Ambiguities compound to generate enormous numbers of possible interpretations.
- In English, a sentence ending in n prepositional phrases has *over* 2^n syntactic interpretations (cf. Catalan numbers).
  - "I saw the man with the telescope": 2 parses
  - "I saw the man on the hill with the telescope.": 5 parses
  - "I saw the man on the hill in Texas with the telescope": 14 parses
  - "I saw the man on the hill in Texas with the telescope at noon.": 42 parses
  - "I saw the man on the hill in Texas with the telescope at noon on Monday" 132 parses

Humor and Ambiguity

- Many jokes rely on the ambiguity of language:
    - Groucho Marx: One morning I shot an elephant in my pajamas. How he got into my pajamas, I'll never know.
    - She criticized my apartment, so I knocked her flat.
    - Noah took all of the animals on the ark in pairs. Except the worms, they came in apples.
    - Policeman to little boy: "We are looking for a thief with a bicycle." Little boy: "Wouldn't you be better using your eyes."
    - Why is the teacher wearing sun-glasses. Because the class is so bright.

NLP Tasks

Syntactic Tasks
Semantic Tasks
Pragmatic Tasks

Syntactic Tasks

❖ **Word Segmentation**
- ❖ identification of boundaries between words

❖ **Morphological Analysis**
- ❖ determination of the smallest meaningful parts of words (morphemes)
  - ❖ e.g. stem + ending

❖ **Part of Speech (PoS) Tagging**
- ❖ annotation of words with labels that describe its role in a sentence

❖ **Phrase Chunking**
- ❖ identification of (non-recursive)
  - ❖ noun phrases (NP)
  - ❖ verb phrases (VP)

❖ **Syntactic Parsing**
- ❖ generation of a syntactically correct parse tree
  - ❖ there may be more than one => ambiguity

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Semantic Tasks

- ❖ **Word Sense Disambiguation (WSD)**
  - ❖ many words in natural languages have multiple possible meaning
  - ❖ WSD tries to determine the most appropriate one
    - ❖ usually requires an interpretation (understanding) of the word and its context

- ❖ **Semantic Role Labeling (SRL)**
  - ❖ analyses the relationship of nouns to the verb in a clause (partial sentence)
    - ❖ e.g., subject, object, source, destination

- ❖ **Semantic Parsing**
  - ❖ mapping of a natural-language sentence to a semantic representation
    - ❖ consistent with the original sentence
    - ❖ correct (according to a semantic specification mechanism
    - ❖ complete (no information omitted)

- ❖ **Textual Entailment**
  - ❖ does one natural-language sentence entail another sentence?
    - ❖ logical inference
    - ❖ ideally: interpretation intended by the speaker/author of the sentence

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN·FH
MÜNCHEN

Textual Entailment Problems from PASCAL Challenge

TEXT	HYPOTHESIS	ENTAILMENT
Eyeing the huge market potential, currently led by Google, Yahoo took over search company Overture Services Inc last year.	*Yahoo bought Overture.*	TRUE
Microsoft's rival Sun Microsystems Inc. bought Star Office last month and plans to boost its development as a Web-based device running over the Net on personal computers and Internet appliances.	*Microsoft bought Star Office.*	FALSE
The National Institute for Psychobiology in Israel was established in May 1971 as the Israel Center for Psychobiology by Prof. Joel.	*Israel was established in May 1971.*	FALSE
Since its formation in 1948, Israel fought many wars with neighboring Arab countries.	*Israel was established in 1948.*	TRUE

Pragmatic Tasks

❖ **Anaphora Resolution**

   ❖ **which phrases in a document refer to the same entity**

      ❖ **in particular he/she/it, this, that, …**

❖ **Ellipsis Resolution**

   ❖ **inferring omitted words from the context**

Related Tasks

Information Extraction
Question Answering
Text Summarization
Machine Translation
Knowledge Acquisition

Information Extraction

❖ **identification of phrases that carry meaningful information**

❖ **Named Entity Recognition (NER)**
  ❖ identification of names of people, places, organizations
  ❖ relies mostly on nouns
  ❖ in English, capitaliziation is an important hint
  ❖ not so useful in German since all nouns are capitalized

❖ **Relation Extraction**
  ❖ identifies relations between entities in the text
  ❖ relies mostly on verbs

❖ **often done off-line for sets of documents that don't change too much**
  ❖ can be combined with search engine activities

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Question Answering

❖ **answers natural-language questions based on information from a set of documents (corpus)**

❖ **relies on information extraction**

46

Text Summarization

❖ **creation of a short summary from a longer piece of text**
  ❖ consistent
    ❖ same meaning as the original
  ❖ correct
    ❖ properly formed sentences in the natural language
  ❖ partial information
    ❖ details may be omitted
    ❖ it can be difficult to determine which details to omit

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN·FH
MÜNCHEN

Machine Translation

❖ **automatic translation of a sentence from one language to another**
  ❖ consistent
    ❖ identical interpretations for both sentences
    ❖ can be difficult to achieve due to semantic and pragmatic differences
      ❖ especially for poetry, puns and jokes, proper etiquette, etc.
  ❖ correct
    ❖ properly formed according to the syntax of the target language
  ❖ complete information
    ❖ nothing should be omitted

❖ **requires ambiguity resolution**
  ❖ syntactic and semantic levels

❖ **pragmatic aspects**
  ❖ usage and conventions of the target language should be respected
    ❖ common phrases and expressions
    ❖ social conventions
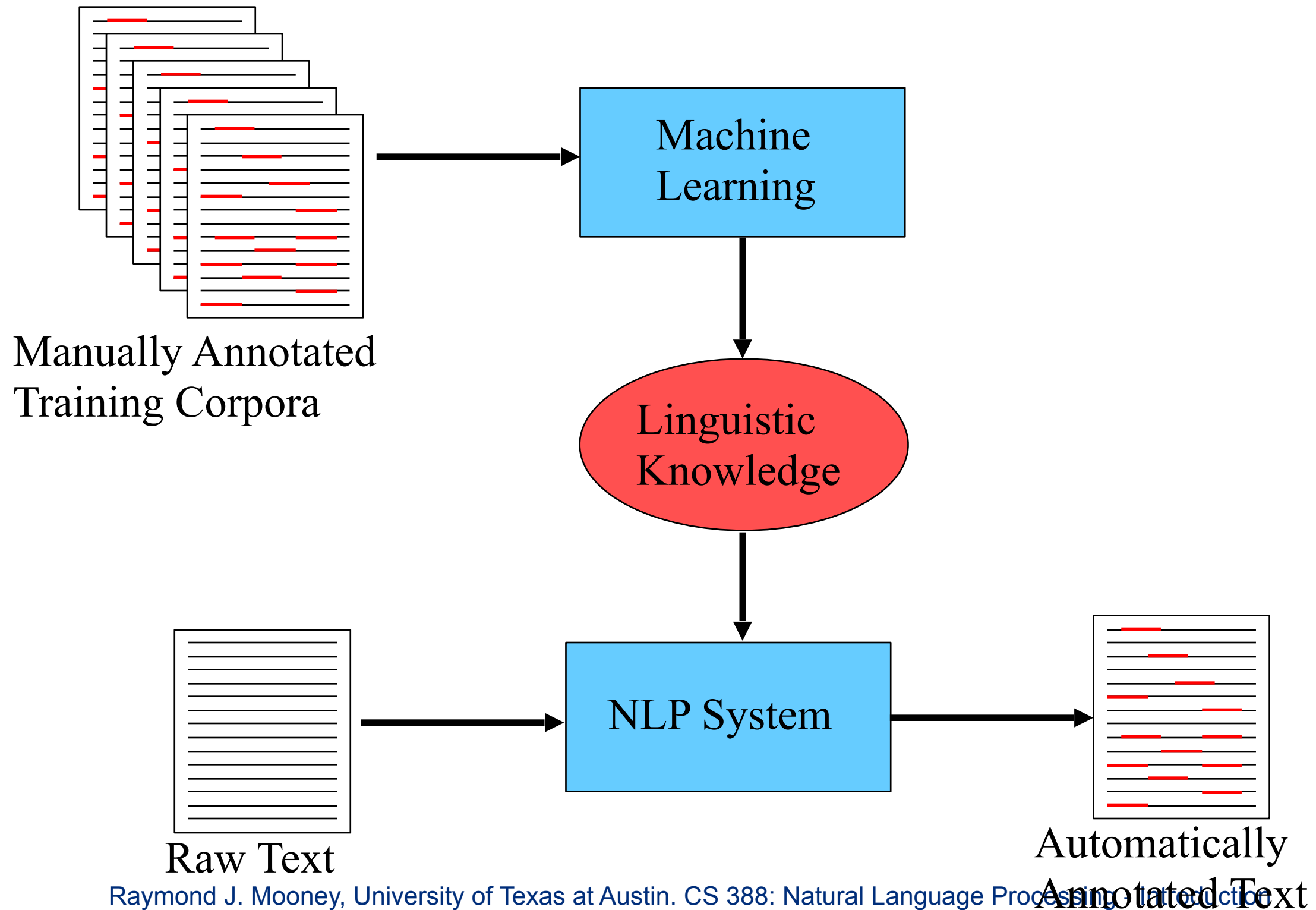   ❖ Du/Sie in German doesn't have a direct counterpart in English

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Translation Mishaps

❖ **translation of English to Russian and then back to English:**

   ❖ "The spirit is willing but the flesh is weak."
     ⇒ "The liquor is good but the meat is spoiled."

   ❖ "Out of sight, out of mind."
     ⇒ "Invisible idiot."

CAL POLY

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Automated Knowledge Acquisition

❖ **use of machine learning and information extraction to populate knowledge bases**

❖ **strong reliance on statistical methods since the 1990's**

Machine Learning Approach

Manually Annotated
Training Corpora

Machine
Learning

Linguistic
Knowledge

Raw Text

NLP System

Automatically
Annotated Text

Deep Learning Approach

- ❖ **relies on large data sets and complex learning mechanisms**
  - ❖ extraction of hierarchically arranged features

- ❖ **avoids the need for manually annotated training data**

- ❖ **may better reflect properties of the original data set (corpus)**

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN

Important Concepts and Terms

Chapter Summary

Sources

❖ **Raymond J. Mooney, University of Texas at Austin. CS 388: Natural Language Processing - Introduction**

❖ **Richard Socher, Stanford University. CS224d: Deep Learning for Natural Language Processing**

CAL POLY

55

HOCHSCHULE
FÜR ANGEWANDTE
WISSENSCHAFTEN · FH
MÜNCHEN