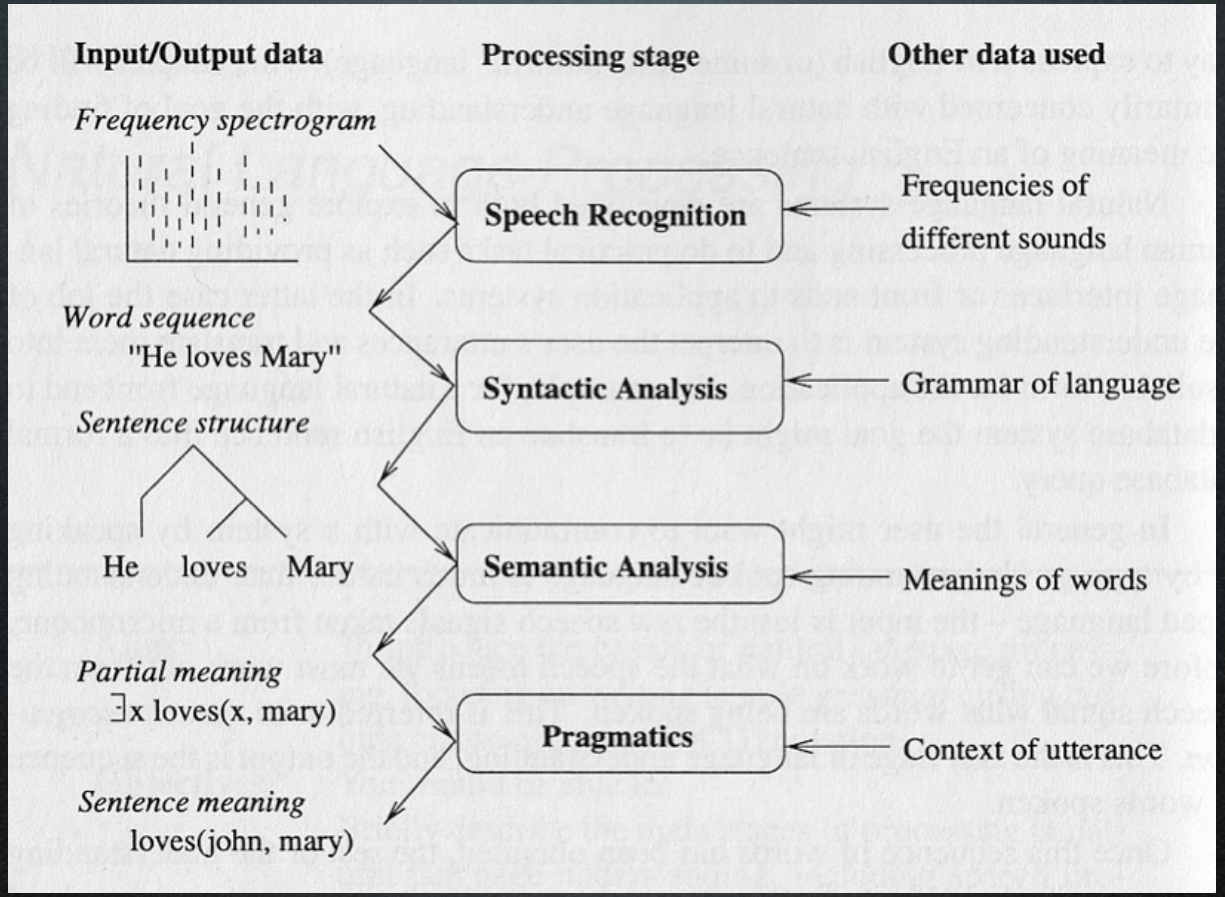
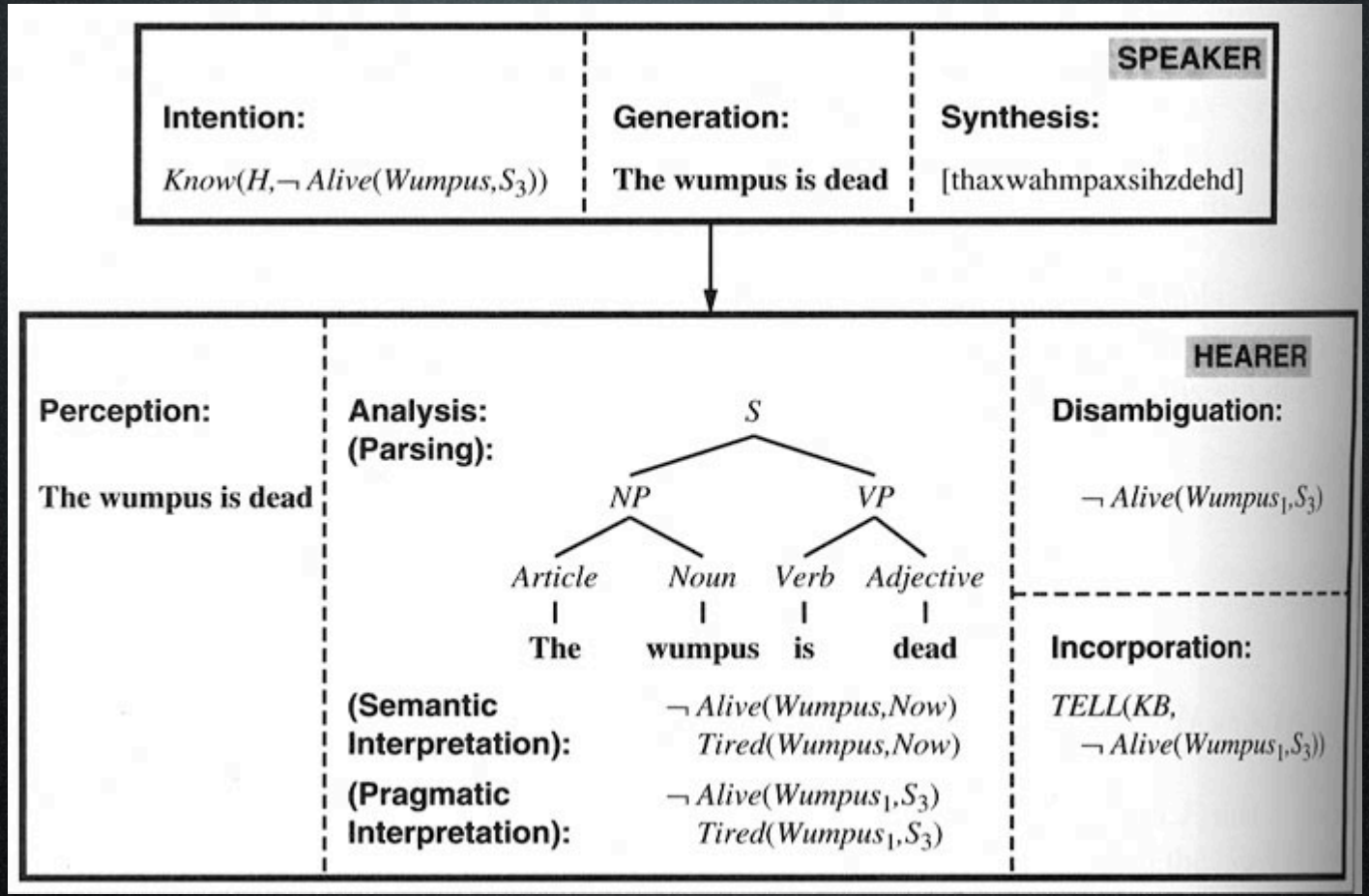
# More of the Same Boring AI Stuff

M. Anthony Kapolka III
CS 340 AI Fall 2019
Wilkes University

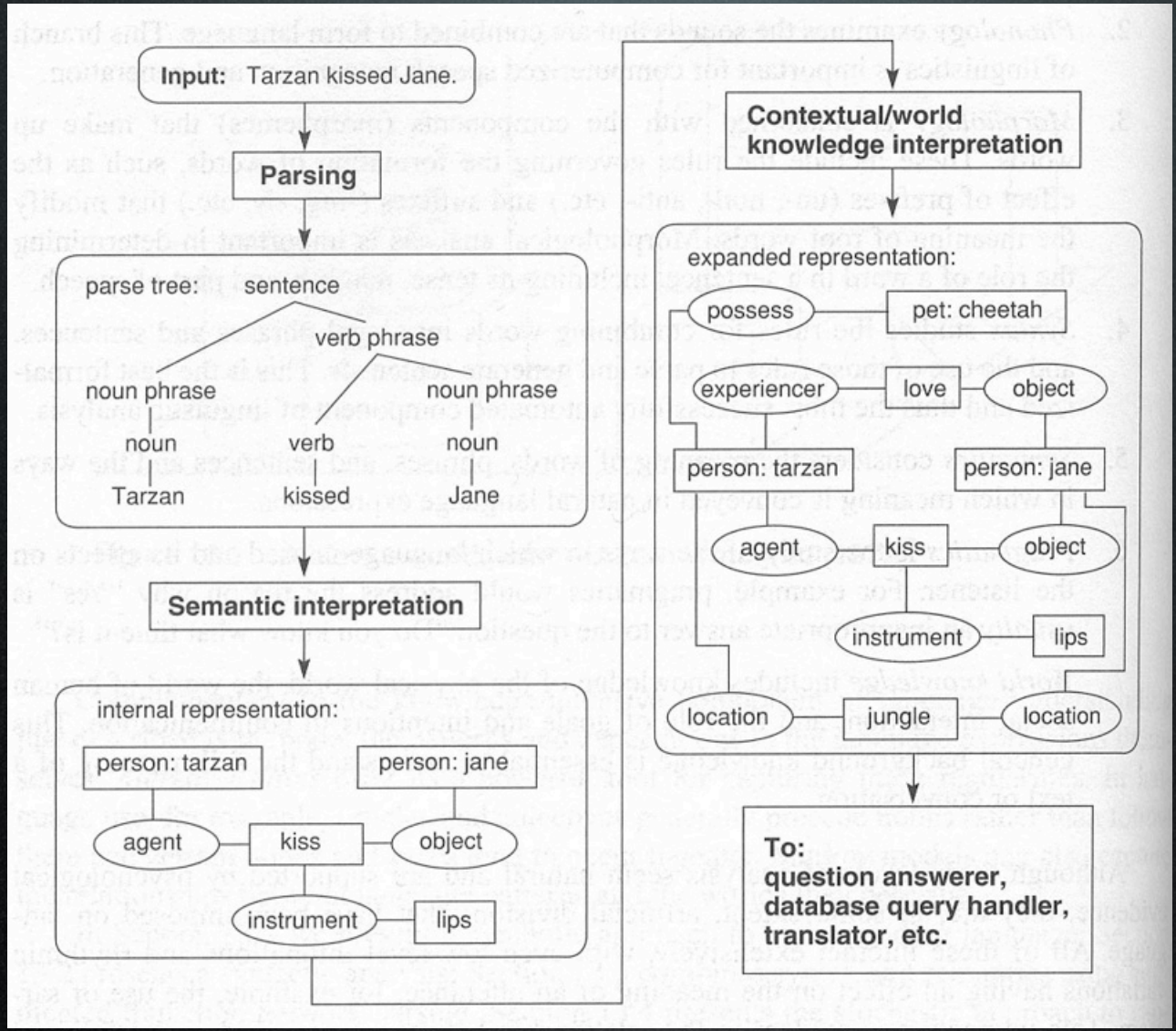
How to wreck a nice beach.



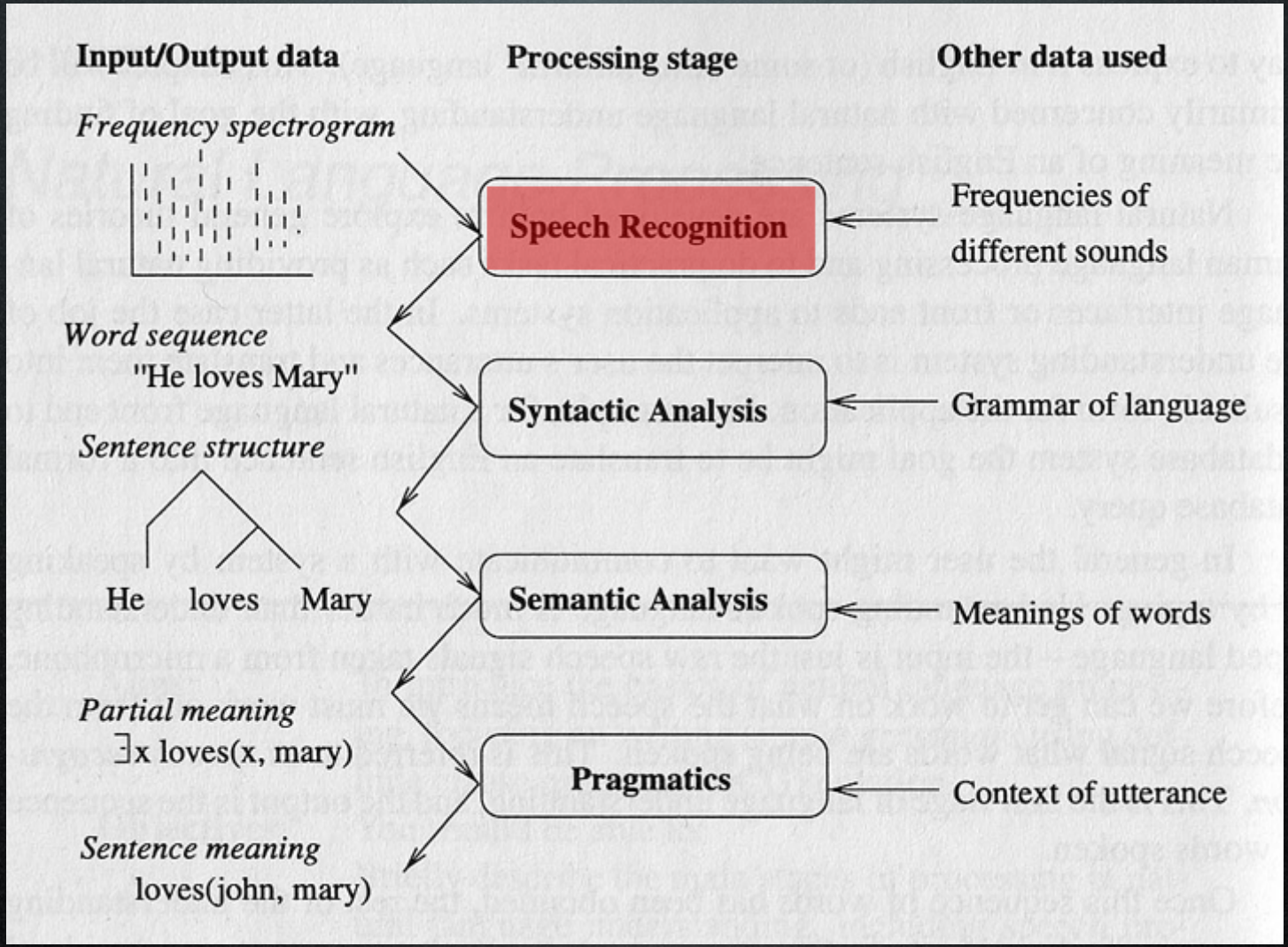

# How to recognize speech.


SPEAKER
Intention:
Know(H,¬ Alive(Wumpus,S3))
Generation:
The wumpus is dead
Synthesis:
[thaxwahmpaxsihzdehd]
HEARER
Perception:
The wumpus is dead
Analysis:
(Parsing):
S
NP
VP
Article
Noun
Verb
Adjective
The
wumpus
is
dead
(Semantic
Interpretation):
¬ Alive(Wumpus,Now)
Tired(Wumpus,Now)
(Pragmatic
Interpretation):
¬ Alive(Wumpus1,S3)
Tired(Wumpus1,S3)
Disambiguation:
¬ Alive(Wumpus1,S3)
Incorporation:
TELL(KB,
¬ Alive(Wumpus1,S3))

[R&N 2003]

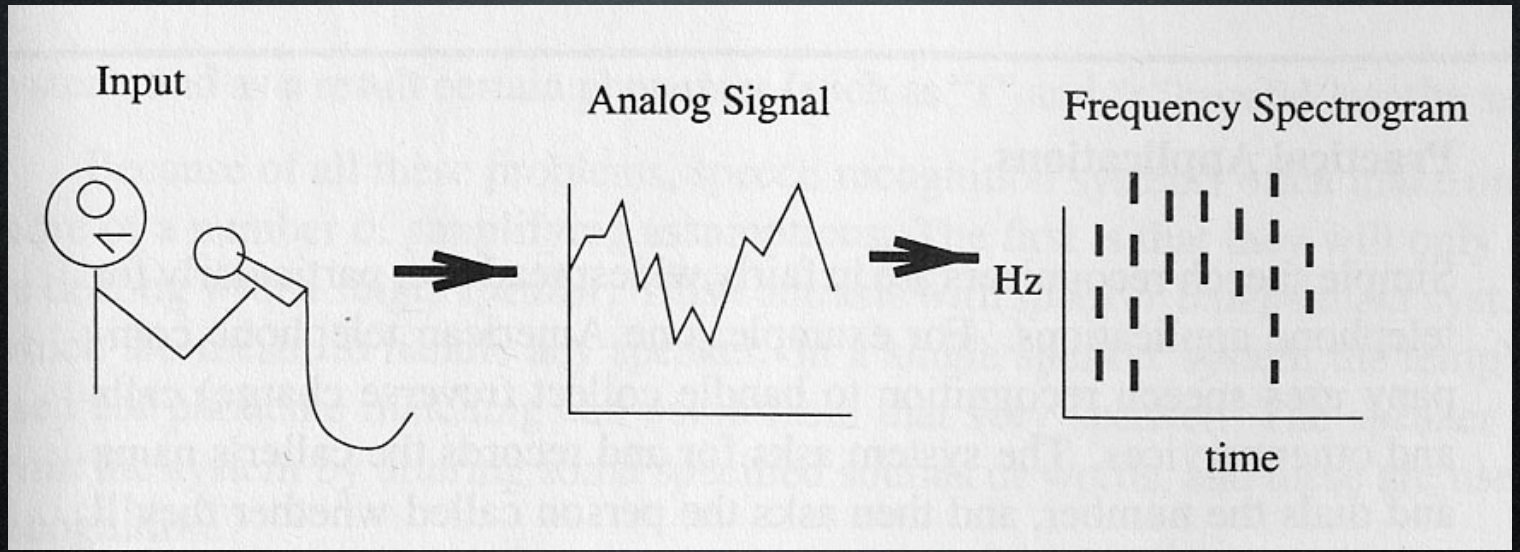

Input: Tarzan kissed Jane.
Parsing
parse tree:
sentence
verb phrase
noun phrase
noun phrase
noun
verb
noun
Tarzan
kissed
Jane
Semantic interpretation
internal representation:
person: tarzan
person: jane
agent
kiss
object
instrument
lips
Contextual/world knowledge interpretation
expanded representation:
possess
pet: cheetah
experiencer
love
object
person: tarzan
person: jane
agent
kiss
object
instrument
lips
location
jungle
location
To:
question answerer,
database query handler,
translator, etc.

[Lugar 2002]


Input/Output data
Processing stage
Other data used
Frequency spectrogram
Speech Recognition
Frequencies of different sounds
Word sequence
"He loves Mary"
Syntactic Analysis
Grammar of language
Sentence structure
He loves Mary
Semantic Analysis
Meanings of words
Partial meaning
∃x loves(x, mary)
Pragmatics
Context of utterance
Sentence meaning
loves(john, mary)

[Cawsey 1997]

# Signal Processing




[Cawsey 1997]

# Signal Processing

- Sampling rate
  - how often a data point is recorded
  - typically 8-16 KHz
- Quantization (Precision)
  - Number of bits for each sample point
  - typically 8-12 bits

# Lots of Data

8 KHz, 8 bit sampling =
about 1/2 MB per minute

# Speech Recognition

Mapping of digitally encoded acoustic signal into a string of characters

(e.g. SPEECH -> TEXT.)

# Speech Comprehension

- Goal: ability to understand meaning
- Harder than recognition  - domain-specific
- Easier than recognition - close enough is ok
- Can use context to identify words

# Speech Recognition

- Constrained
  - many commercial vendors, good success
  - lightweight apps running on PDAs & cellphones

# constrained voice recognition

WED 30 NOV 2005

## Get Your Own R2

READ MORE: R2-D2, ROBOTS

I know you've been waiting for this—don't deny it. Finally, R2-D2 can be yours, in your own home, and for only $119.95. Fully functional, R2 can roam around and obeys over 40 voice commands (please refrain from asking him to project Obi-Wan please). He also plays children's games like tag with an IR sensor that can be set to detect motion (which can also be set to sound an alarm for a watchdog effect). Available at Hammacher Schlemmer.

Star Wars R2-D2 Robot Replica On Sale [i4u]

R2-D2 Robot Product Page [Hammacher Schlemmer]

# Speech Recognition

- Unconstrained (e.g. Dictation)
  - few vendors, commercial systems
  - constrained by dictionary
  - adaptive, requires training
  - error rates 1 order of magnitude > human
  - most users disappointed
  - human-like performance projected 2050.

# Speech Recognition

- Human speech consists of 40-50 phonemes
- A phoneme is a basic speech sound,
- Some letters/letter combinations map to multiple phonemes.

# English Phonemes

| | | | |
|---|---|---|---|
| [iy] , beat | [ih] , bit | [ey] , bet | [ae] , bat |
| [ah] , but | [ao] , bought | [ow] , boat | [uh] , book |
| [ux] , beauty | [er] , Bert | [ay] , buy | [oy] , boy |
| [axr] , diner | [aw] , down | [ax] , about | [ix] , roses |
| [aa] , cot | [b] , bet | [ch] , Chet | [d] , debt |
| [f] , fat | [g] , get | [hh] , hat | [hv] , high |
| [jh] , jet | [k] , kick | [l] , let | [el] , bottle |
| [m] , met | [em] , bottom | [n] , net | [en] , button |
| [ng] , sing | [eng] , Washington | [p] , pet | [r] , rat |
| [s] , set | [sh] , shoe | [t] , ten | [th] , thick |
| [dh] , that | [dx] , butter | [v] , vet | [w] , wet |
| [wh] , which | [y] , yet | [z] , zoo | [zh] , measure |

# Detecting Phonemes

- Often the signal is processed with a sliding window (called a frame)
- typical width 10 milliseconds

# Process signal...

- Identify features (frequency, amplitude) to identify phonemes.
- Using dictionary, assemble phonemes into words

# Process signal...

- Identify features (frequency, amplitude) to identify phonemes.
- Using dictionary, assemble phonemes into words

[k], [ae], [t] ☞ cat

# Problems

- homophones (homonyms)
- segmentation (identifying word breaks - not always present in fluent speech)

# Solutions

- Extract most likely string of words
- Extract multiple possible words
- Invoke Parser / Semantic Analysis

# Can use a probabilistic model

This model is broken down into components using Bayes rule.

$$P(\text{words} \mid \text{signal}) = (P(\text{words}) \times P(\text{signal} \mid \text{words}))/P(\text{signal})$$

Given a signal, the task is to find the sequence of words that maximizes P(words I signal).

(P(words)×P(signal | words))/P(signal)

P(signal | words) is the acoustic model.

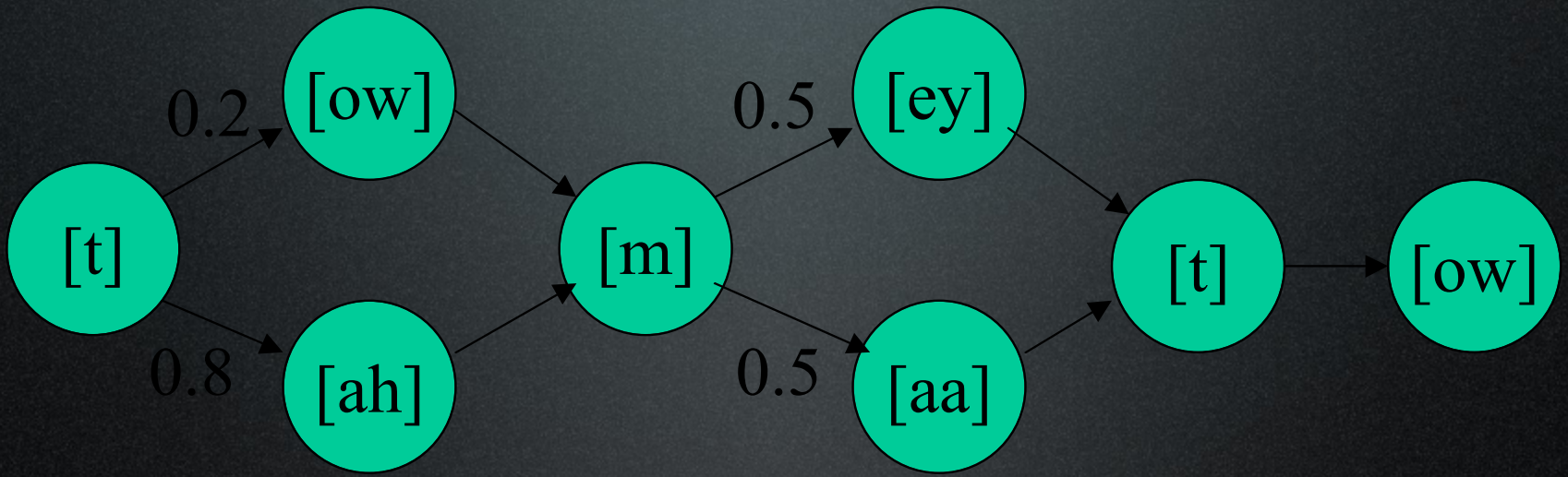
[k], [ae], [t] ☞ cat

Problems:

1) dialects (different pronunciations)
2) coarticulation (quick, slurred speech)

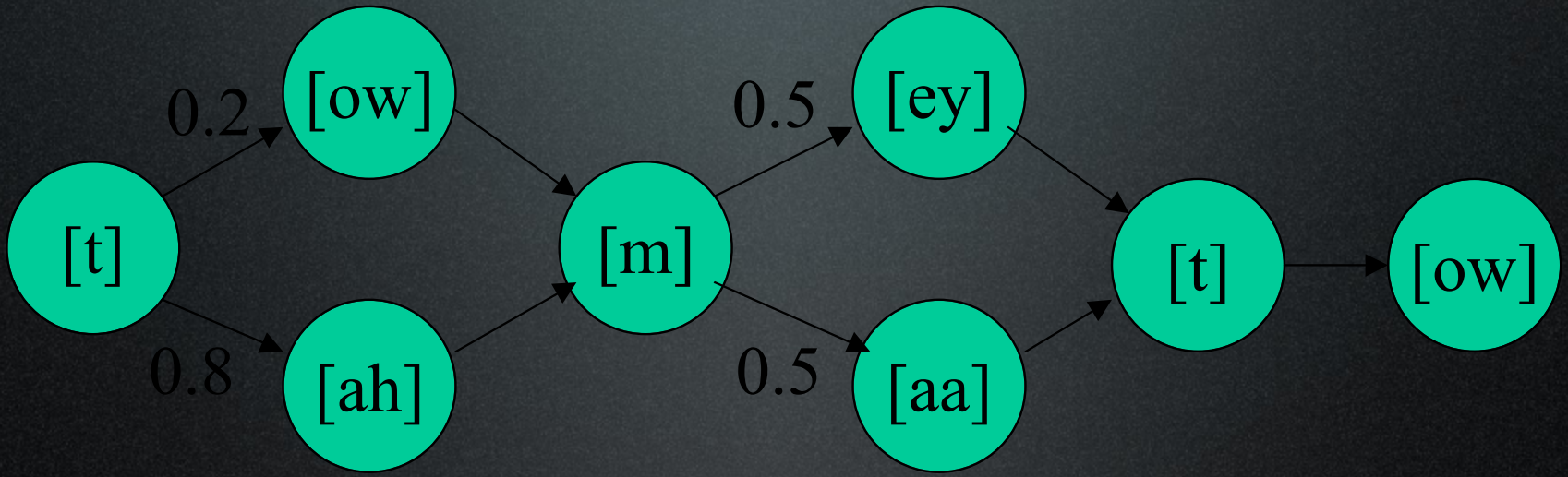

Josh Diak, Citizen of Australia

# you say tomato...




example of a markov model showing all phoneme combinations for tomato (dialect & coarticulation)

# you say tomato...




P([towmeytow] | “tomato”) = 0.1
P([tahmaatow] | “tomato”) = 0.4

I say...

t ah mey t ae hh

$$(P(words) \times P(signal \mid words)) / P(signal)$$

P(words) is the language model.

That is, probability based
on language context.

(P(words)×P(signal | words))/P(signal)

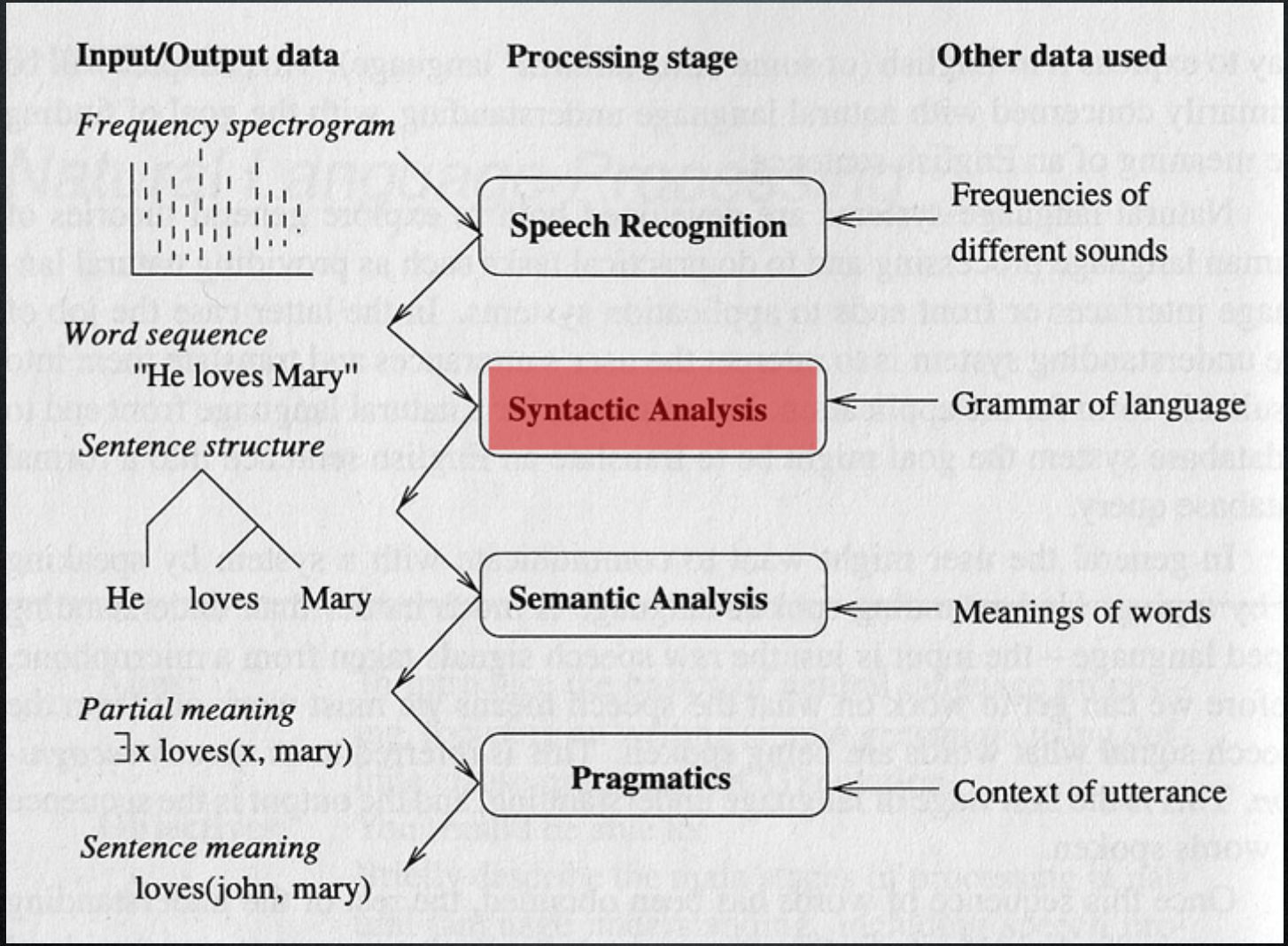
the man is saying “I have a …”

(P(words)×P(signal I words))/P(signal)

P(signal | gum) = 0.7  P(signal | gun) = 0.4

$$(P(words) \times P(signal \mid words))/P(signal)$$

- P(words) is a factor to add weight to more likely utterances.
- How to compute P(words) ??
- Coincident words
- Feedback from speech comprehension.


Input/Output data
Processing stage
Other data used
Frequency spectrogram
Speech Recognition
Frequencies of different sounds
Word sequence
"He loves Mary"
Syntactic Analysis
Grammar of language
Sentence structure
He loves Mary
Semantic Analysis
Meanings of words
Partial meaning
∃x loves(x, mary)
Pragmatics
Context of utterance
Sentence meaning
loves(john, mary)

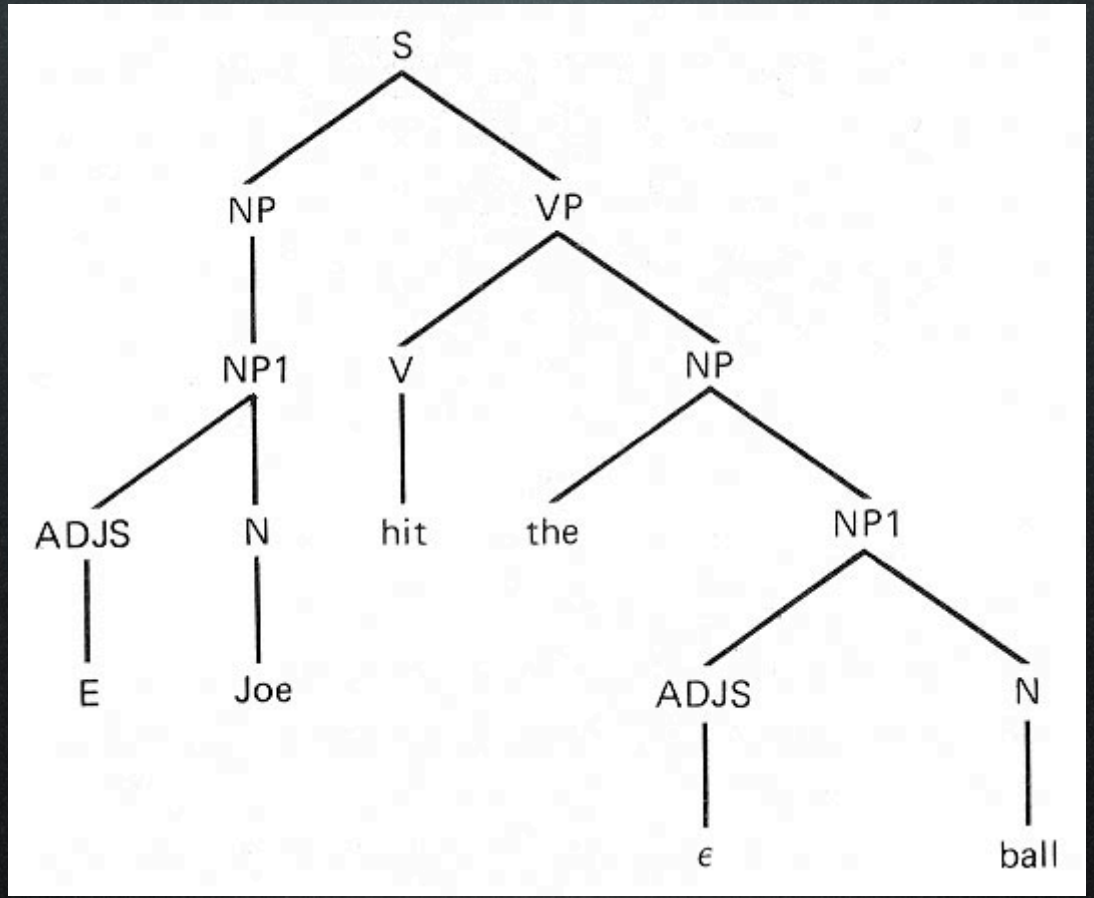
[Cawsey 1997]

# Grammar

S → NP VP
NP → the NP1
NP → NP1
NP1 → ADJS N
ADJS → ϵ | ADJ ADJS
VP → V
VP → V NP
N → Joe | boy | ball
ADJ → little | big
V → hit | ran

[Rich 1991]

# Parse Tree



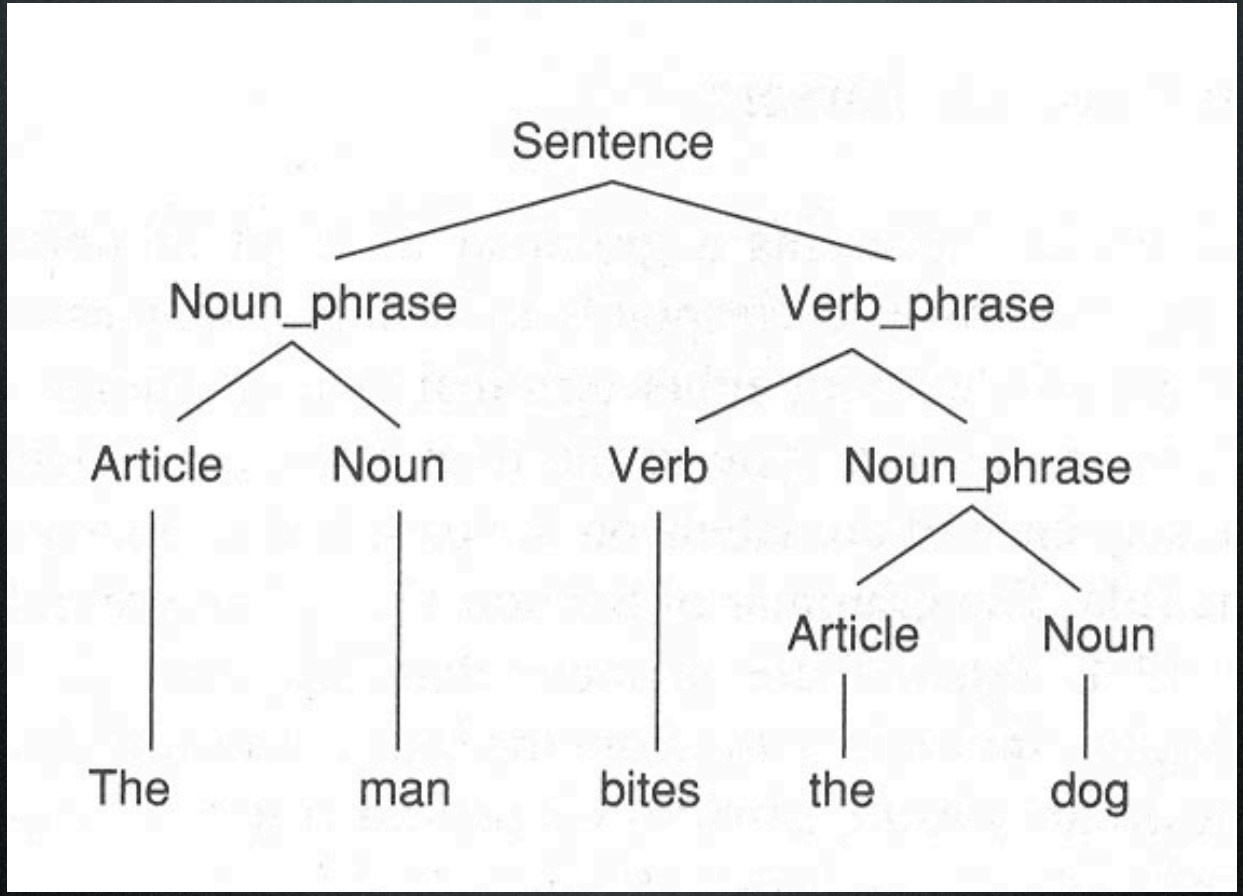

[Rich 1991]

# A Dog Grammar

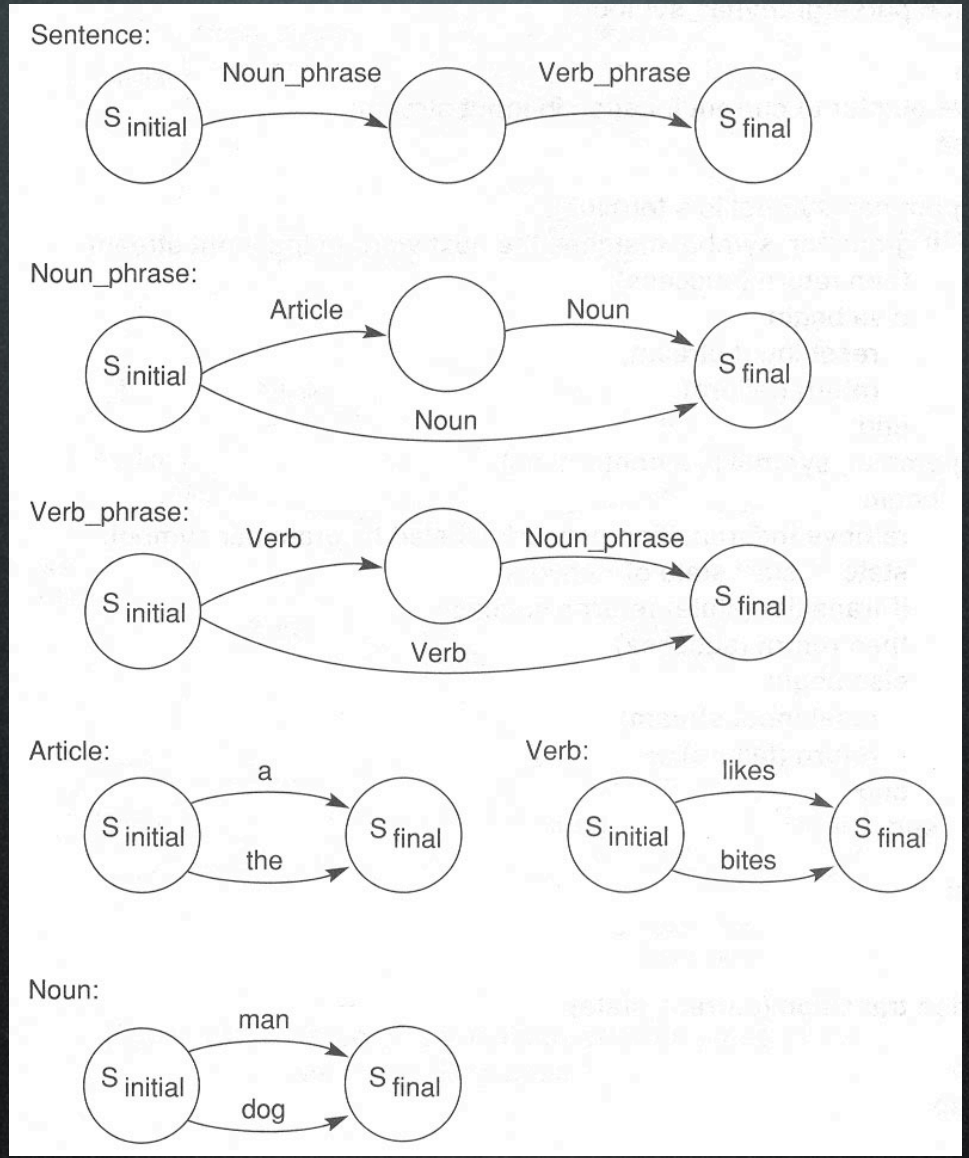
1. sentence ↔ noun_phrase verb_phrase
2. noun_phrase ↔ noun
3. noun_phrase ↔ article noun
4. verb_phrase ↔ verb
5. verb_phrase ↔ verb noun_phrase
6. article ↔ a
7. article ↔ the
8. noun ↔ man
9. noun ↔ dog
10. verb ↔ likes
11. verb ↔ bites

[Lugar 2002]

# And Parse Tree

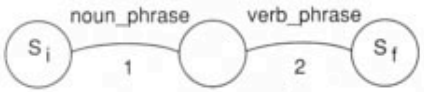



[Lugar 2002]


Sentence:
Noun_phrase
Verb_phrase
S initial
S final
Noun_phrase:
Article
Noun
S initial
S final
Noun
Verb_phrase:
Verb
Noun_phrase
S initial
S final
Verb
Article:
a
S initial
S final
the
Verb:
likes
S initial
S final
bites
Noun:
man
S initial
S final
dog

[Lugar 2002]

sentence:



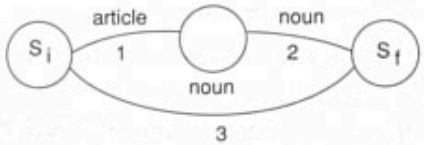

```
function sentence-1;
  begin
    NOUN_PHRASE := structure returned by
            noun_phrase network;

    SENTENCE.SUBJECT := NOUN_PHRASE;
  end.
```

```
function sentence-2;
  begin
    VERB_PHRASE := structure returned by
            verb_phrase network;

     if NOUN_PHRASE.NUMBER =
                                VERB_PHRASE.NUMBER
       then begin
          SENTENCE.VERB_PHRASE := VERB_PHRASE;
          return SENTENCE
        end
        else fail
  end.
```

noun_phrase:



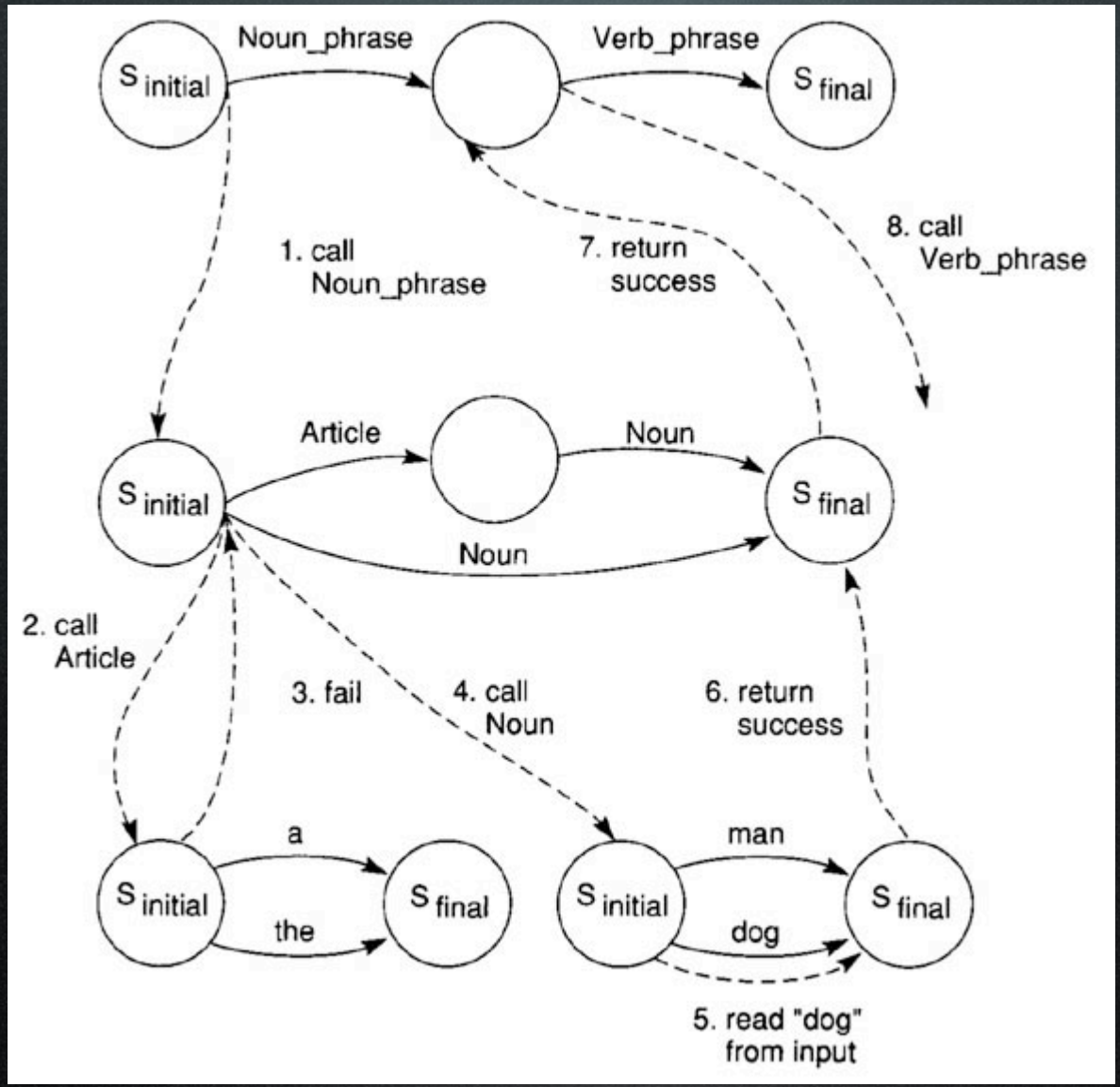

```
function noun_phrase-1;
  begin
    ARTICLE := definition frame for next word of input;

    if ARTICLE.PART_OF_SPEECH=article
      then NOUN_PHRASE.DETERMINER := ARTICLE
      else fail
  end.
```

```
function noun_phrase-2;
  begin
    NOUN := definition frame for next word of input;

    if NOUN.PART_OF_SPEECH=noun and
          NOUN.NUMBER agrees with
                NOUN_PHRASE.DETERMINER.NUMBER
      then begin
        NOUN_PHRASE.NOUN := NOUN;
        NOUN_PHRASE.NUMBER := NOUN.NUMBER
        return NOUN_PHRASE
      end
      else fail
  end.
```

[Lugar 2002]


Noun_phrase
Verb_phrase
S initial
S final
1. call
Noun_phrase
7. return
success
8. call
Verb_phrase
Article
Noun
S initial
S final
Noun
2. call
Article
3. fail
4. call
Noun
6. return
success
a
man
S initial
S final
S initial
S final
the
dog
5. read "dog"
from input

[Lugar 2002]

# Classic Jokes

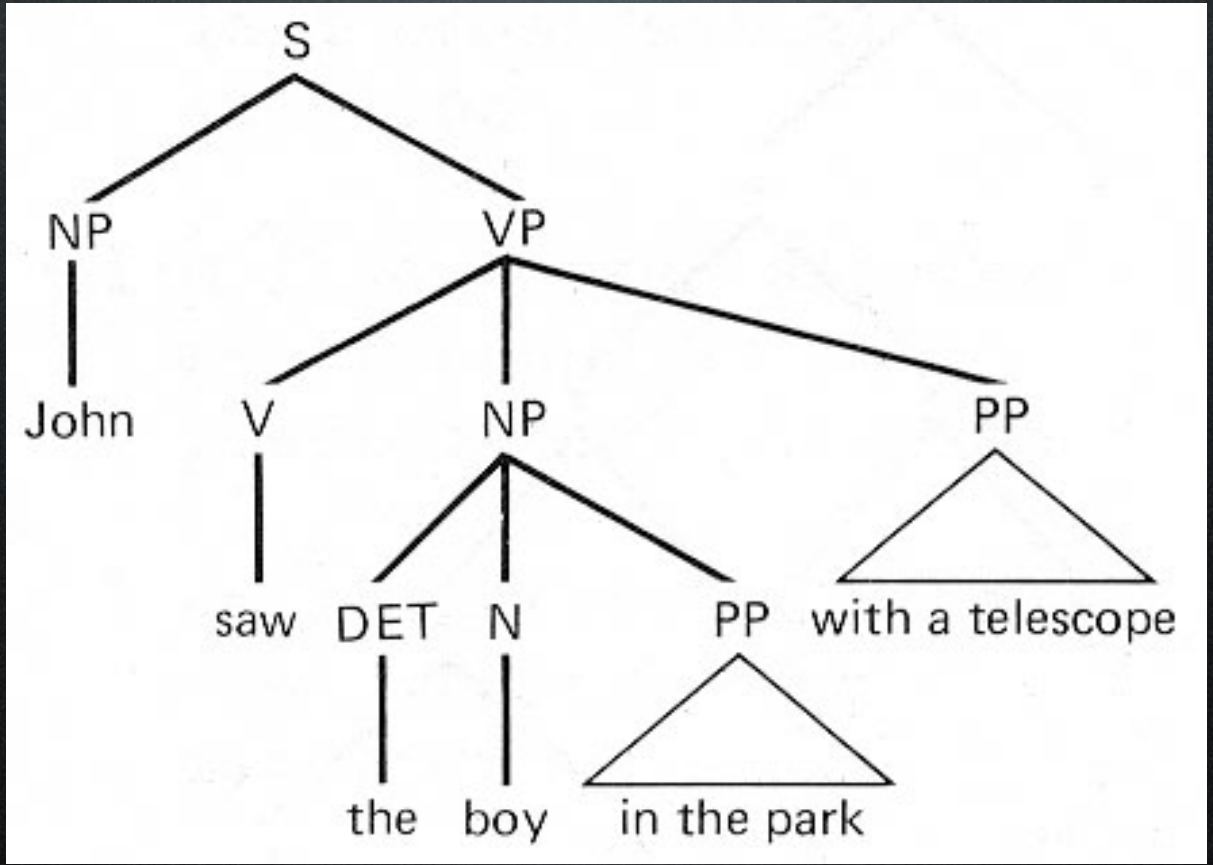
Time flies like an arrow.

# Classic Jokes

Time flies like an arrow.

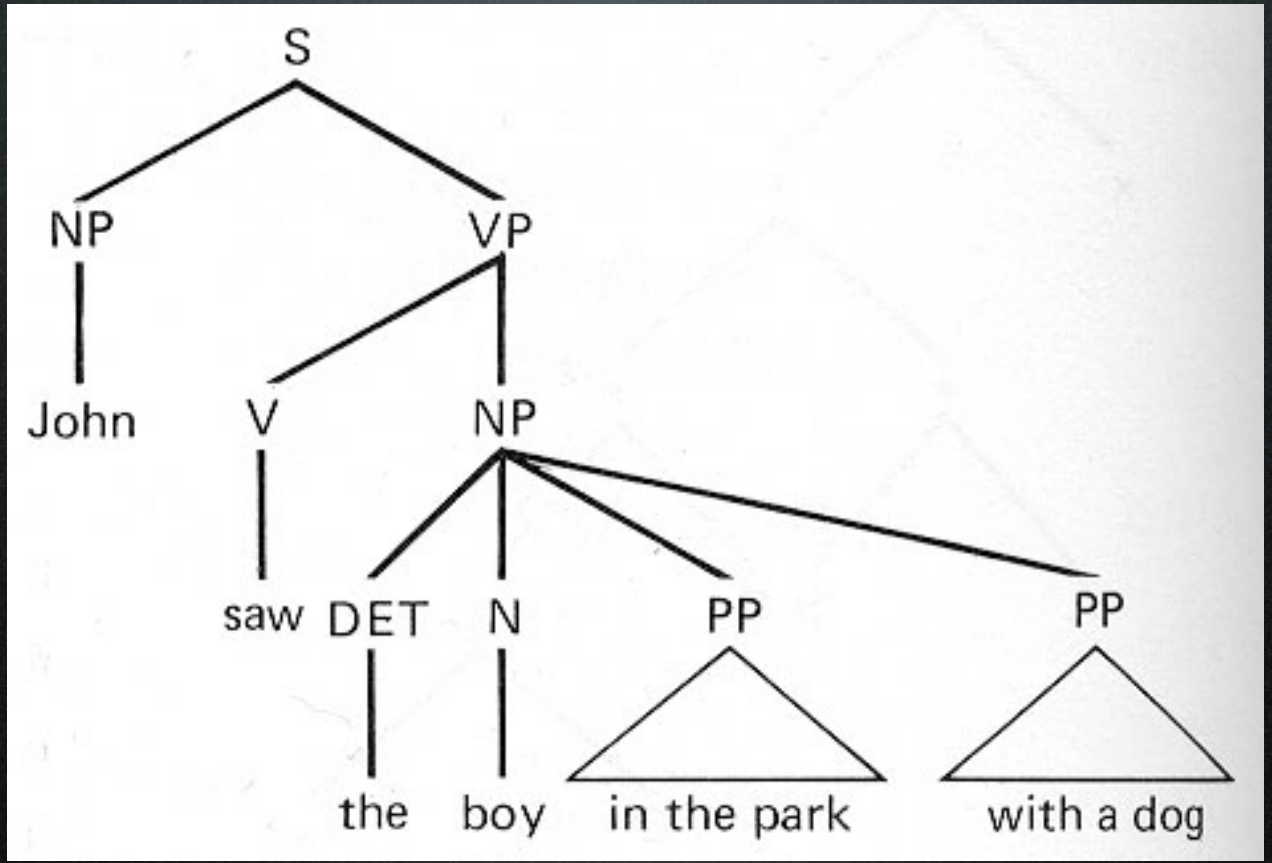
Fruit flies like a banana.

# Syntactic Ambiguity



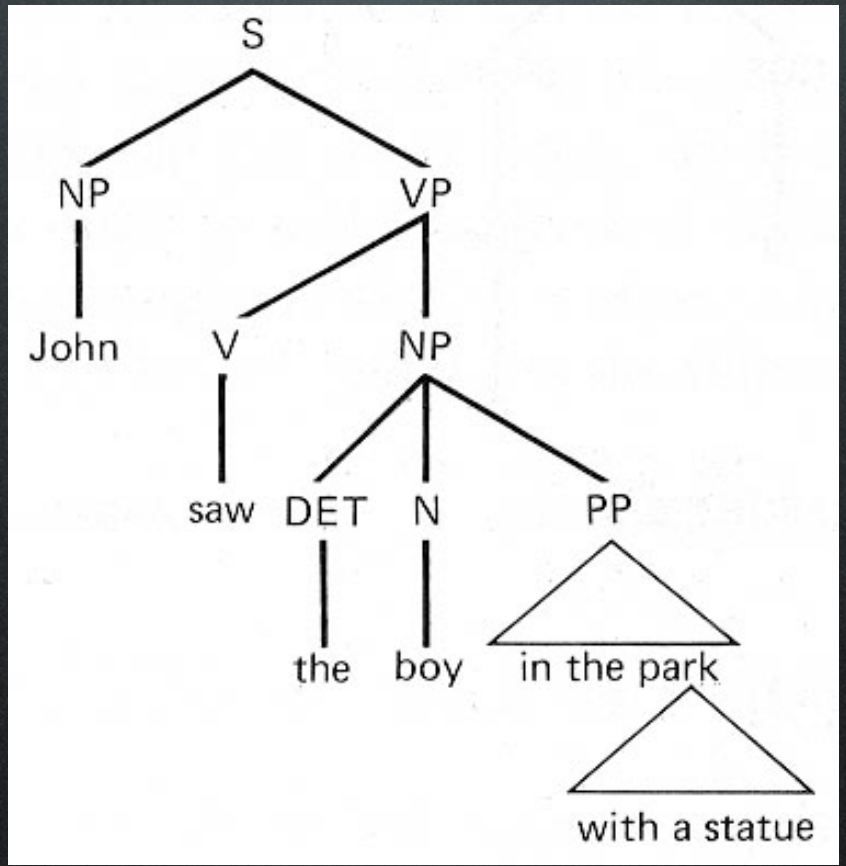

[Rich 1991]

# Syntactic Ambiguity




[Rich 1991]

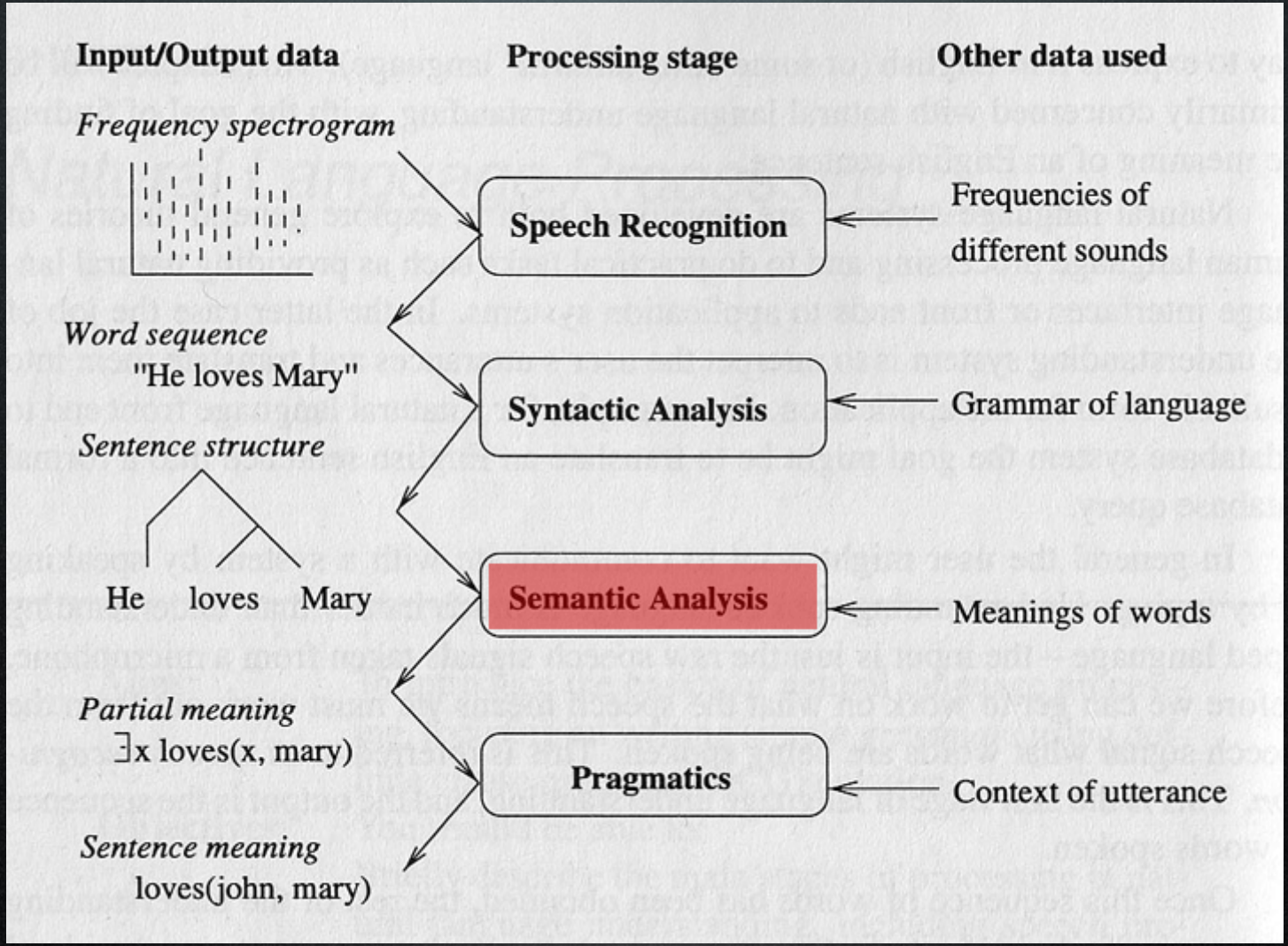
# Syntactic Ambiguity




[Rich 1991]

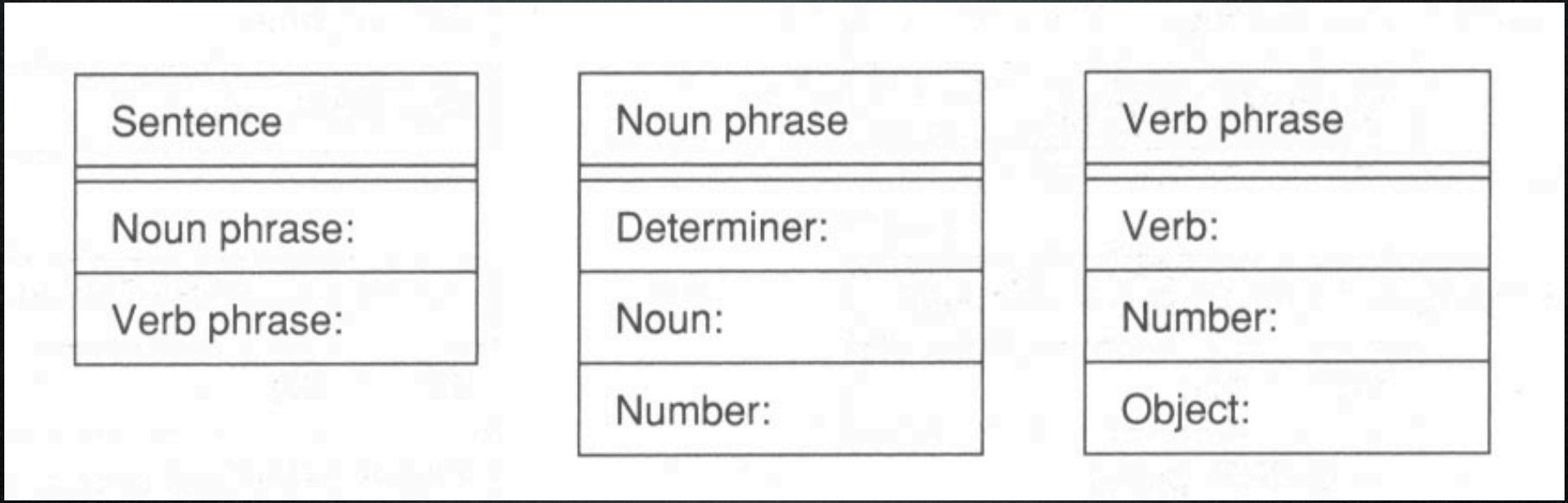
# Syntactic Ambiguity

- How do we disambiguate?
- Context
- Feedback from speech comprehension


Input/Output data
Processing stage
Other data used
Frequency spectrogram
Speech Recognition
Frequencies of different sounds
Word sequence
"He loves Mary"
Syntactic Analysis
Grammar of language
Sentence structure
He loves Mary
Semantic Analysis
Meanings of words
Partial meaning
∃x loves(x, mary)
Pragmatics
Context of utterance
Sentence meaning
loves(john, mary)

[Cawsey 1997]

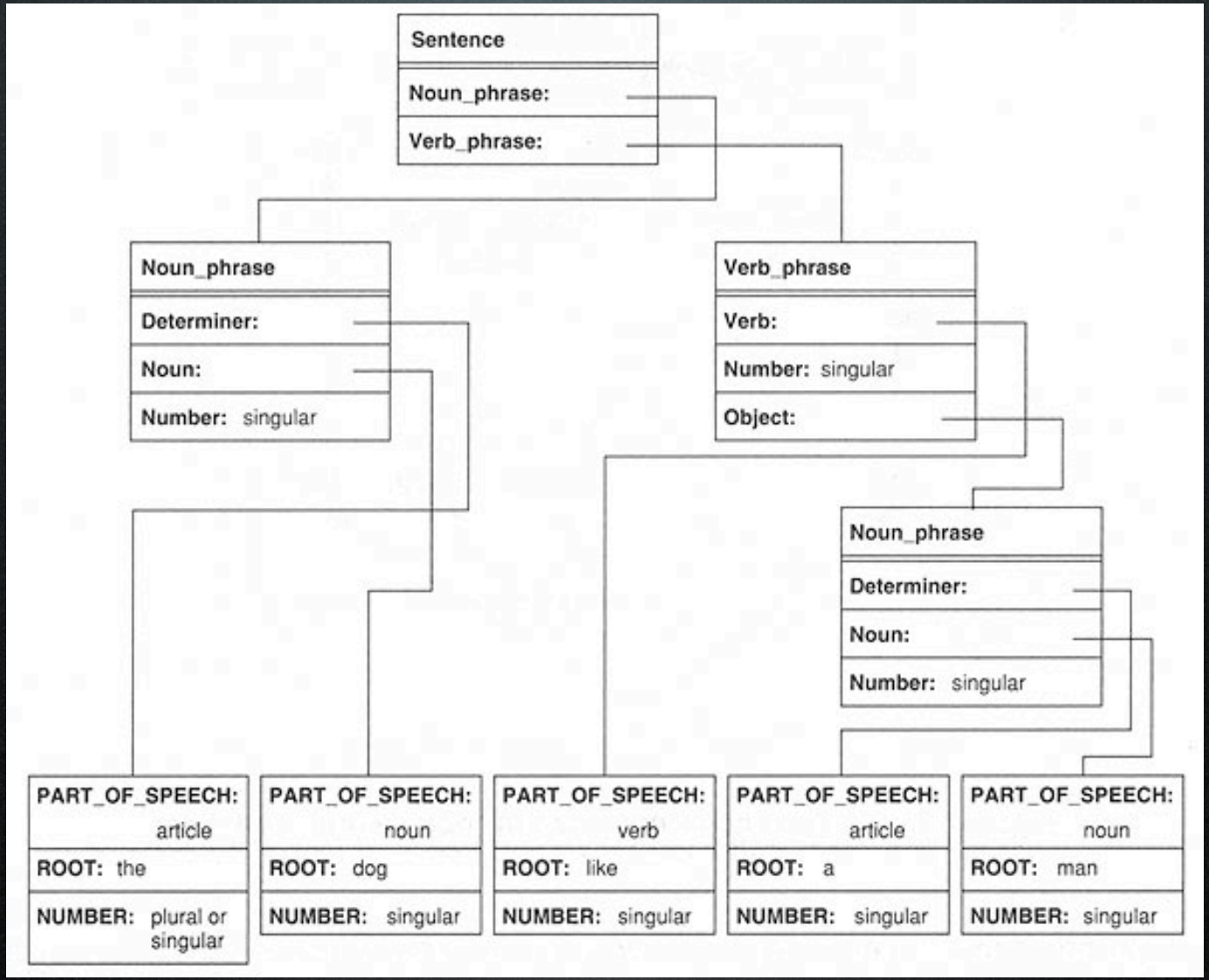
# Can use a slot-filler ds



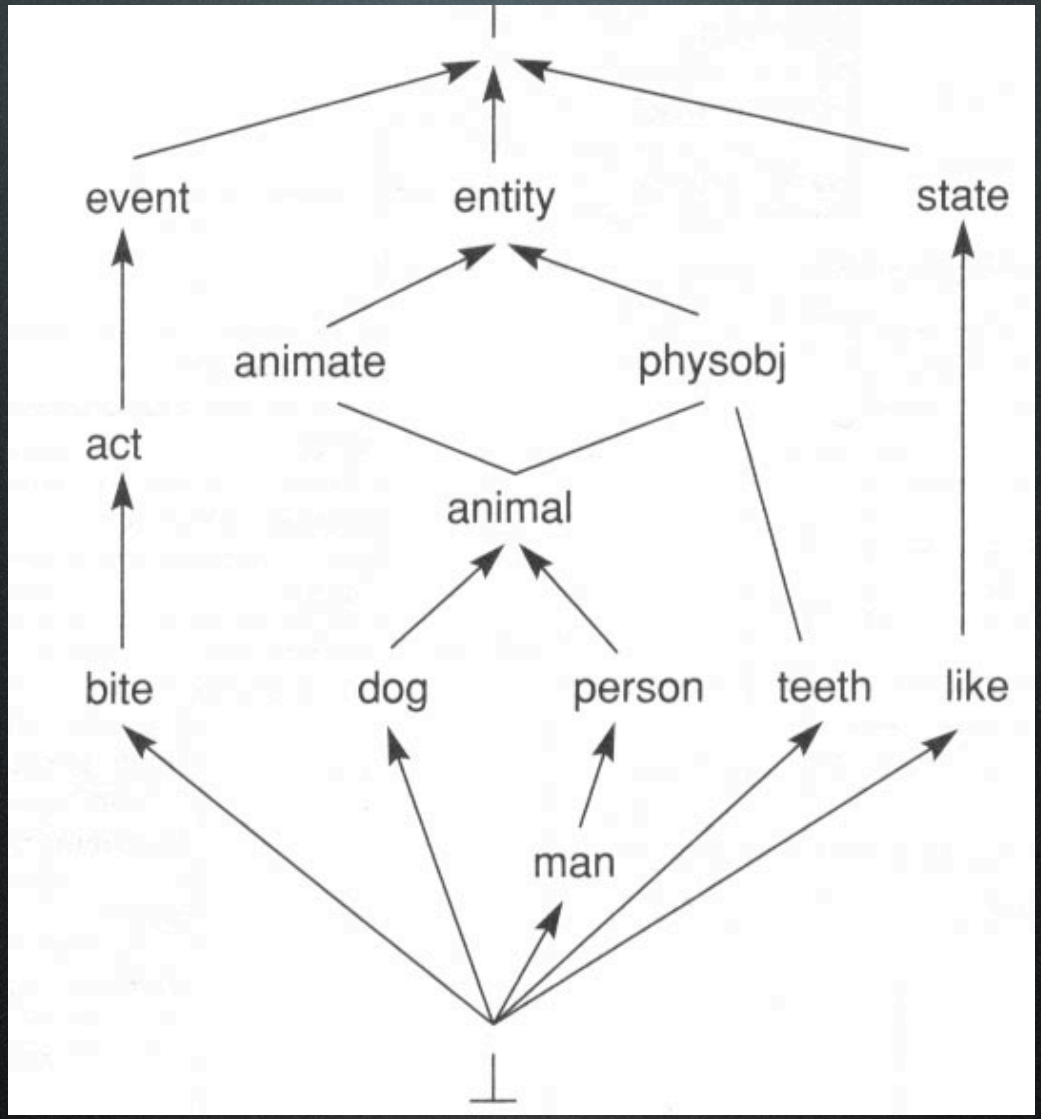

[Lugar 2002]

| Word | Definition | Word | Definition |
|---|---|---|---|
| a | PART_OF_SPEECH: article<br>ROOT: a<br>NUMBER: singular | like | PART_OF_SPEECH: verb<br>ROOT: like<br>NUMBER: plural |
| bite | PART_OF_SPEECH: verb<br>ROOT: bite<br>NUMBER: plural | likes | PART_OF_SPEECH: verb<br>ROOT: like<br>NUMBER: singular |
| bites | PART_OF_SPEECH: verb<br>ROOT: bite<br>NUMBER: singular | man | PART_OF_SPEECH: noun<br>ROOT: man<br>NUMBER: singular |
| dog | PART_OF_SPEECH: noun<br>ROOT: dog<br>NUMBER: singular | men | PART_OF_SPEECH: noun<br>ROOT: man<br>NUMBER: plural |
| dogs | PART_OF_SPEECH: noun<br>ROOT: dog<br>NUMBER: plural | the | PART_OF_SPEECH: article<br>ROOT: the<br>NUMBER: plural or singular |

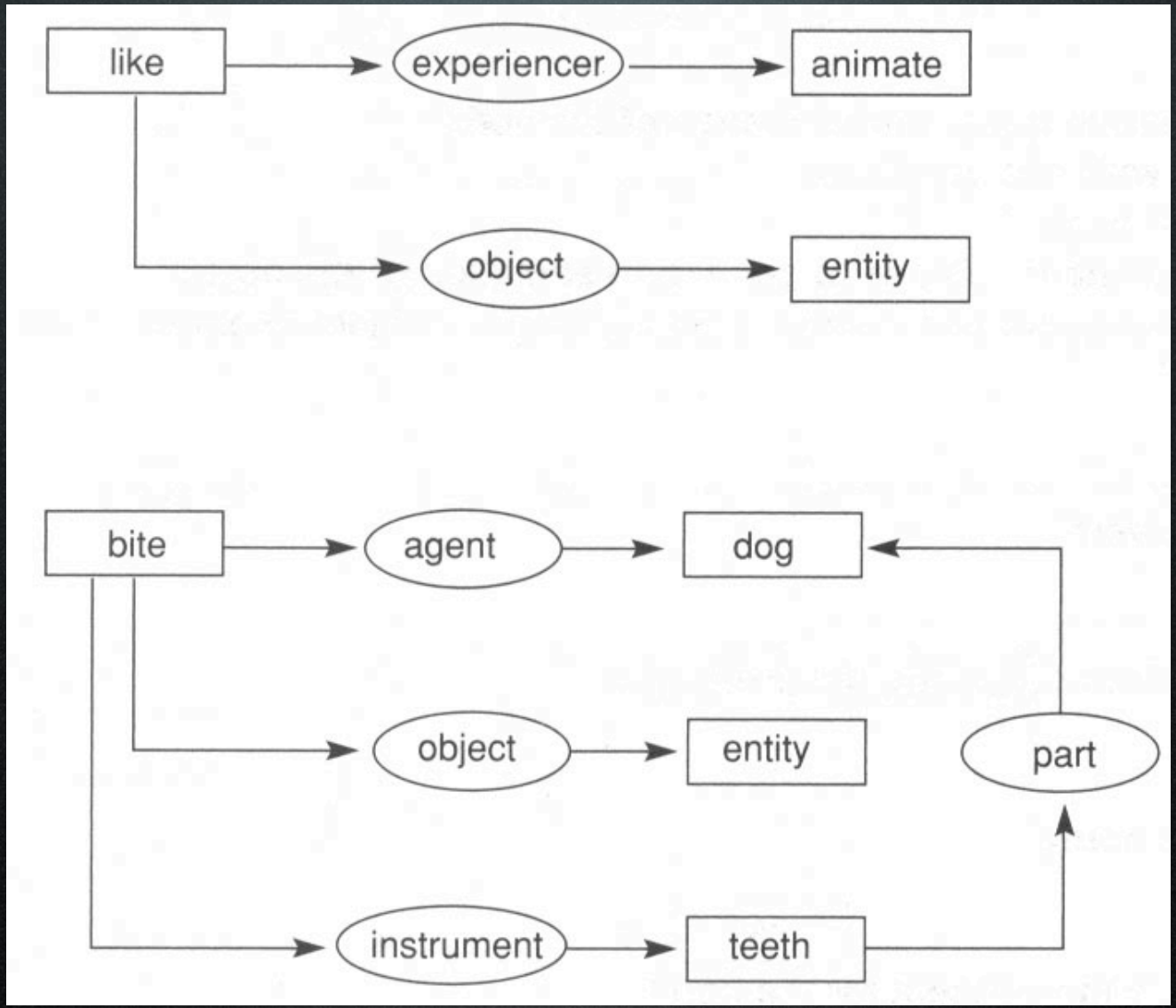
[Lugar 2002]


Sentence
Noun_phrase:
Verb_phrase:
Noun_phrase
Determiner:
Noun:
Number: singular
Verb_phrase
Verb:
Number: singular
Object:
Noun_phrase
Determiner:
Noun:
Number: singular
PART_OF_SPEECH: article
ROOT: the
NUMBER: plural or singular
PART_OF_SPEECH: noun
ROOT: dog
NUMBER: singular
PART_OF_SPEECH: verb
ROOT: like
NUMBER: singular
PART_OF_SPEECH: article
ROOT: a
NUMBER: singular
PART_OF_SPEECH: noun
ROOT: man
NUMBER: singular

[Lugar 2002]

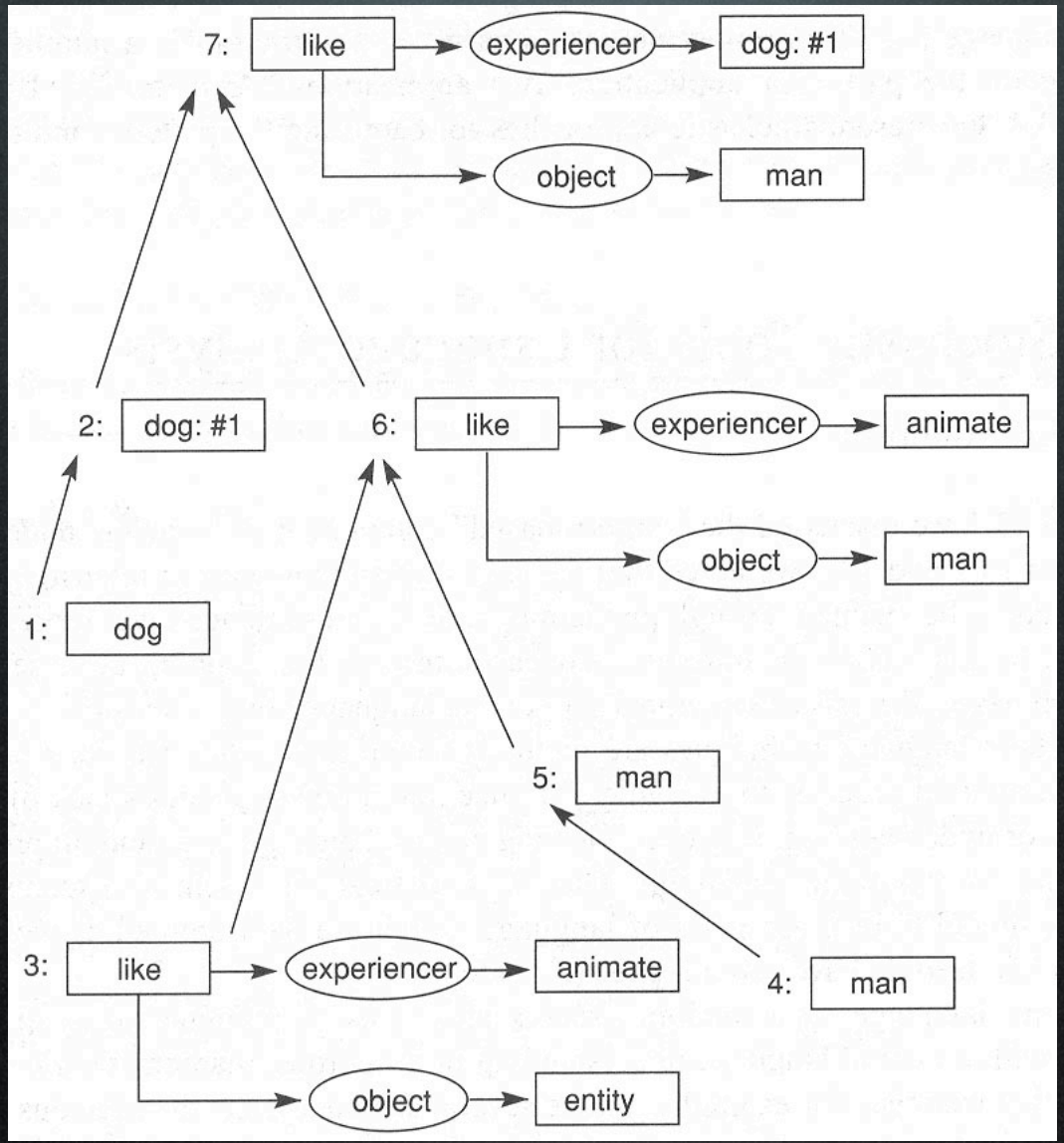

event
entity
state
animate
physobj
act
animal
bite
dog
person
teeth
like
man

[Lugar 2002]

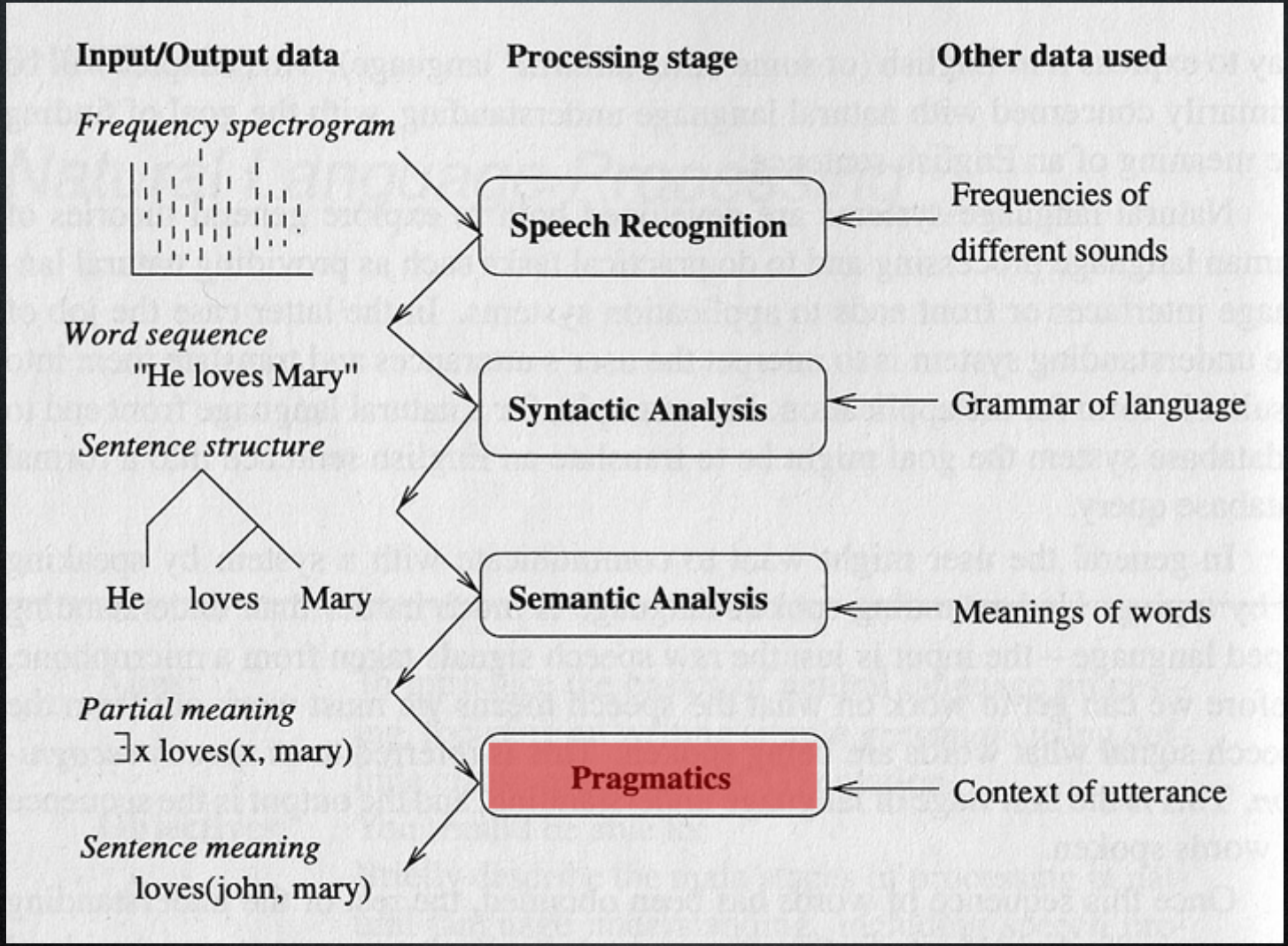

like
experiencer
animate
object
entity
bite
agent
dog
object
entity
part
instrument
teeth

[Lugar 2002]

Augment our parse tree with semantic meaning


7:
like
experiencer
dog: #1
object
man
2:
dog: #1
6:
like
experiencer
animate
object
man
1:
dog
5:
man
3:
like
experiencer
animate
4:
man
object
entity

[Lugar 2002]


Input/Output data
Processing stage
Other data used
Frequency spectrogram
Speech Recognition
Frequencies of different sounds
Word sequence
"He loves Mary"
Syntactic Analysis
Grammar of language
Sentence structure
He loves Mary
Semantic Analysis
Meanings of words
Partial meaning
∃x loves(x, mary)
Pragmatics
Context of utterance
Sentence meaning
loves(john, mary)

[Cawsey 1997]

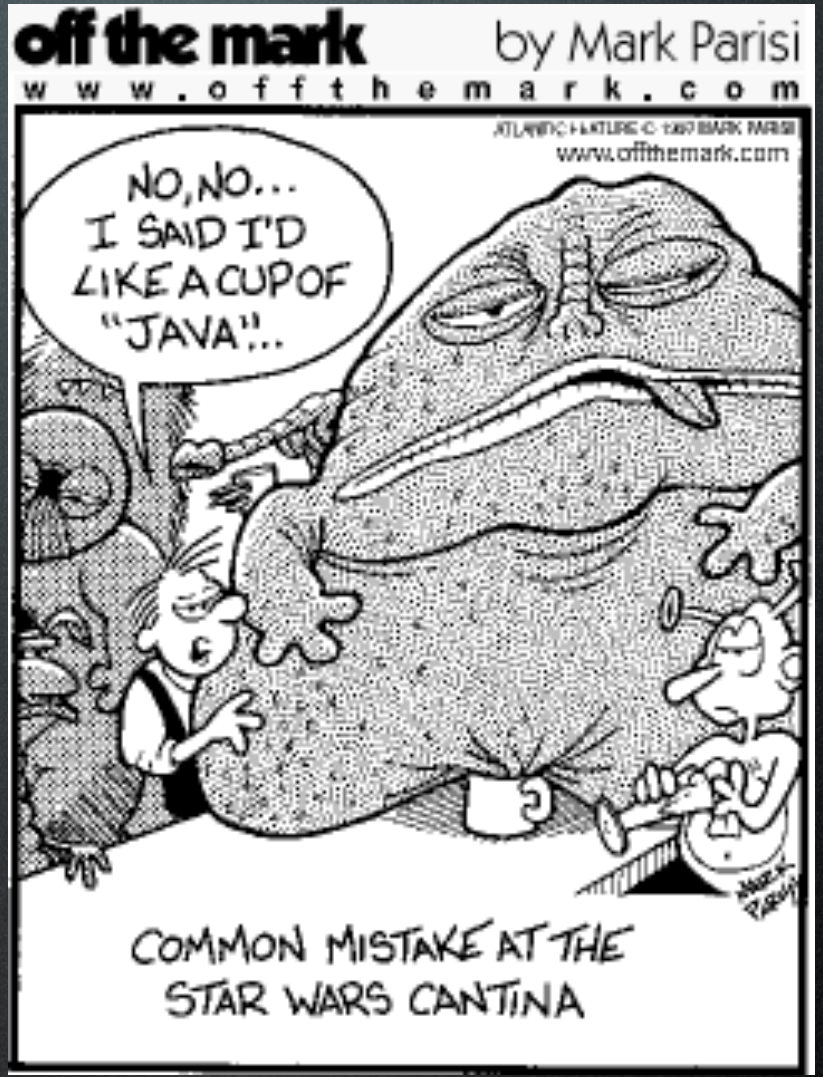
# Pragmatics

Language expresses more than just facts.

Language can convey a deeper purpose.

# When I ask...

Where is the coffee??


off the mark
by Mark Parisi
www.offthemark.com
www.offthemark.com
NO, NO... I SAID I'D LIKE A CUP OF "JAVA"...
COMMON MISTAKE AT THE STAR WARS CANTINA

# When I ask...

## Where is the coffee??


I might actually mean...

I want to make some coffee.

# When I ask...

## Where is the coffee??

I might actually mean...

Would you make some coffee?

# When I ask...

## Where is the coffee??

I might actually mean...

Where did you put my coffee?

Look at the bigger picture

c.f restaurant scripts

# Thematic Roles

Think of context as a scene.

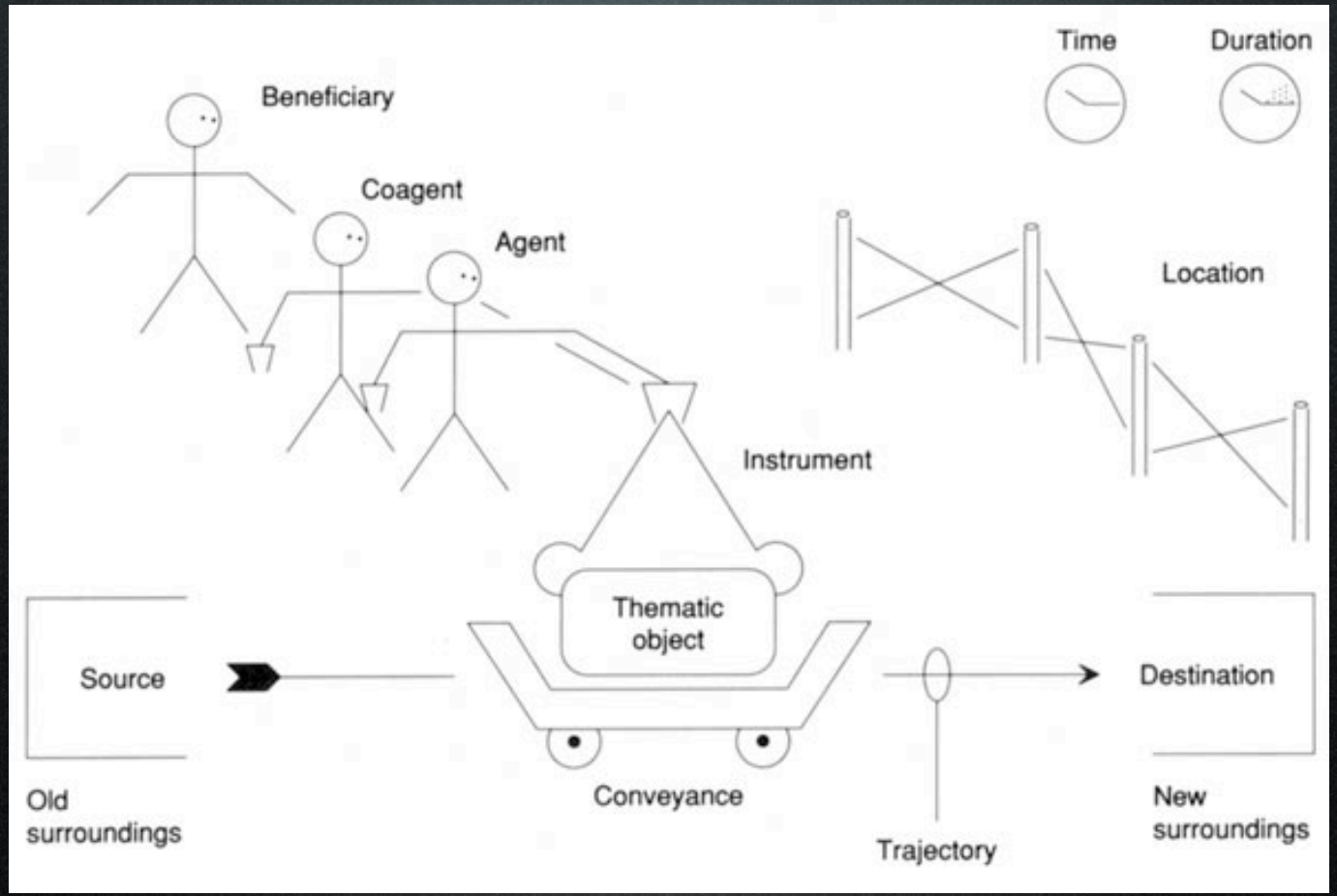
Parts of the scene represented by thematic roles.

Exact number of thematic roles open to debate.

# Example Roles

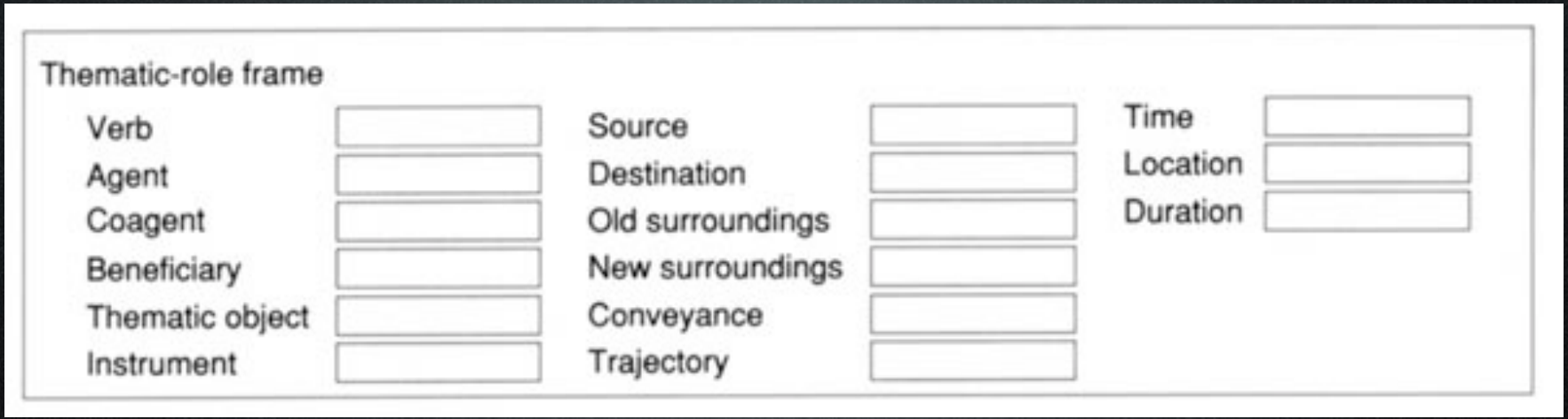
- Agent - causes action to occur
- Coagent - assists causing action
- Beneficiary - for whom action is performed
- Thematic object - thing the sentence is about
- Instrument - tool used by agent

# Thematic Roles




[Winston 1992]

# Thematic Roles




could be implemented as a frame system
roles filled by sentence analysis
carried as context to next sentence
used to build a script, answer questions

[Winston 1992]