COMP474/6741 Natural Language Processing(NLP) and Grammars

Communication

Typical communication episode
S (speaker) wants to convey P (proposition) to H (hearer) using W (words in a formal or natural language)

1. Speaker
- **Intention:** S wants H to believe P
- **Generation:** S chooses words W
- **Synthesis:** S utters words W

2. Hearer
- **Perception:** H perceives words W" (ideally W" = W)
- **Analysis:** H infers possible meanings P1,P2,…,Pn for W"
- **Disambiguation:** H infers that S intended to convey Pi (ideally Pi=P)
- **Incorporation:** H decides to believe or disbelieve Pi

Languages

- Artificial
  - Smaller vocabulary
  - Simple syntactic structures
  - Non-ambiguous
  - Not tolerant to errors (ex. Syntax error)
- Natural
  - Large and open vocabulary (new words everyday)
  - Complex syntactic structures
  - Very ambiguous
  - Robust (ex. forgot a comma, a word... still OK)
  - Evolving

Natural Language Processing (NLP)

1. Natural Language Understanding
   - Taking some spoken/typed sentence and working out what it means
2. Natural Language Generation
   - Taking some formal representation of what you want to say and working out a way to express it in a natural (human) language (e.g., English)

Applications of Nat. Lang. Processing

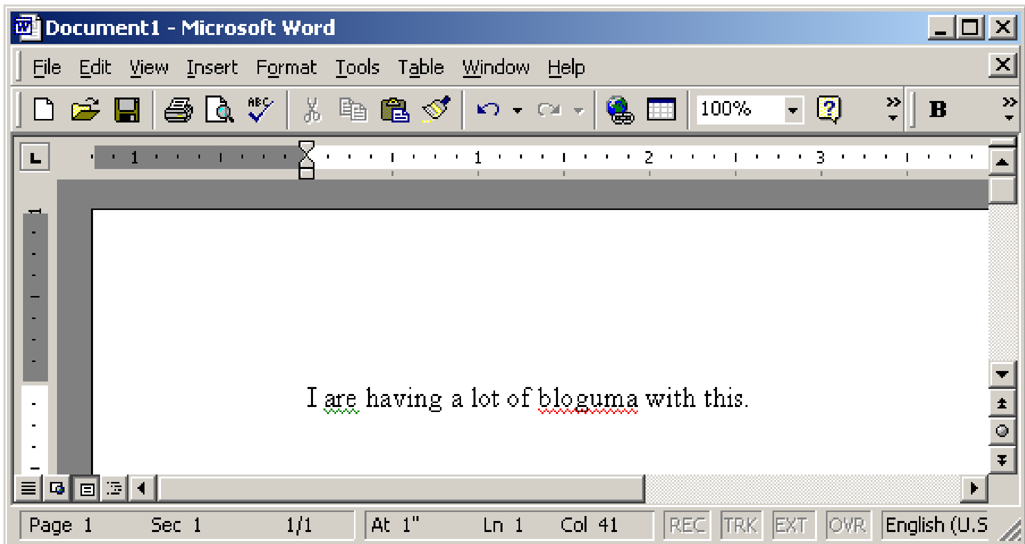
- Machine Translation
- Database Access
- Information Retrieval
  - Selecting from a set of documents the ones that are relevant to a query
- Text Categorization
  - Sorting text into fixed topic categories
- Extracting data from text
  - Converting unstructured text into structure data
- Spoken language control systems
- Spelling and grammar checkers

Practical applications of NLP

- Machine Translation
  - ex: Google translate: http://www.google.com/translate_t
  - ex: Systran: http://www.systransoft.com/index.html
  - ex: **METEO® System**
    translates weather reports from English to French everyday at Environment Canada
- Information Retrieval
  - ex: Google
  - Find documents from a collection that satisfies a user query

Practical applications of NLP

- Spell checker, grammatical corrector



- Text Categorization
  - Sort documents into categories according to their content (ex. sports, politics,…)
- …

Why is NLP hard?

- *The chair's leg is broken.*
- Headline News: *Two sisters reunited after 14 years at the check-out counter.*
- *One morning I shot an elephant in my pajamas. How he got into my pajamas I'll never know.*
- ...

Today

- Full Stages of NLP

1. Speech Recognition
2. Syntax Analysis
3. Semantic Interpretation
    1. Discourse Analysis
4. Pragmatics & World Knowledge

Natural language understanding

Raw speech signal

- **Speech recognition**

Sequence of words spoken

- **Syntactic analysis** using knowledge of the grammar

Structure of the sentence

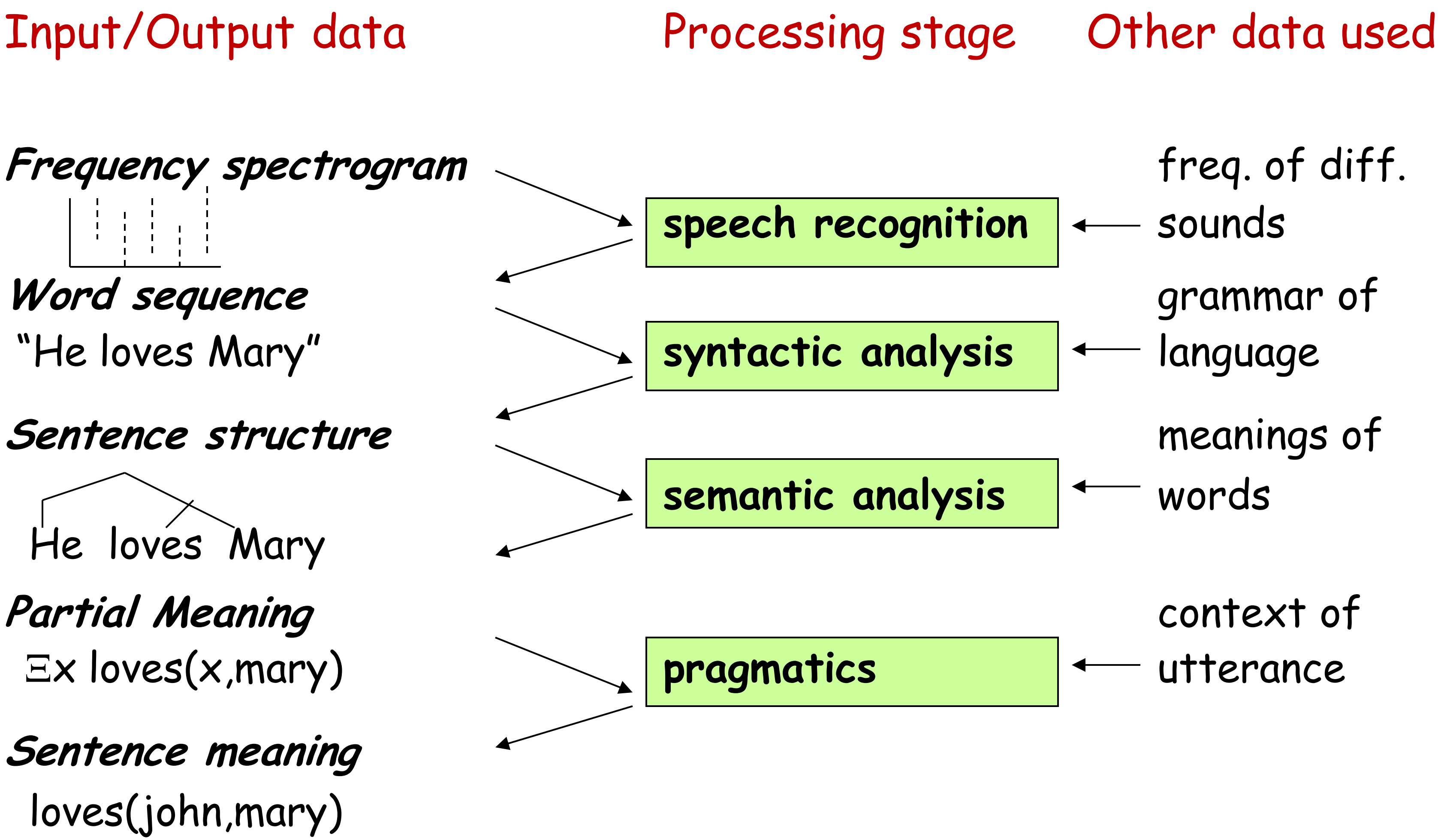
- **Semantic analysis** using info. about meaning of words

Partial representation of meaning of sentence

- **Pragmatic analysis** using info. about context

Final representation of meaning of sentence

Natural Language Understanding



Today

- Full Stages of NLP

1. Speech Recognition

2. Syntax Analysis
3. Semantic Interpretation
    1. Discourse Analysis
4. Pragmatics & World Knowledge

Speech Recognition (1 of 2)

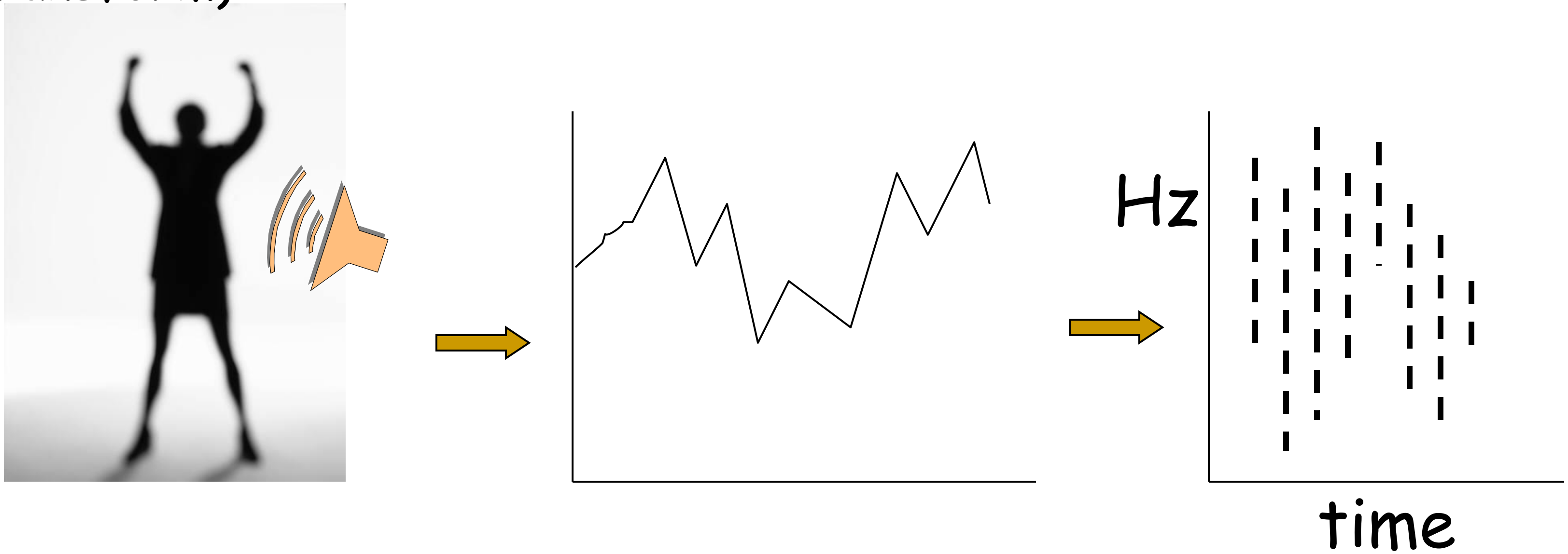
Input

Analog Signal

Freq. spectrogram

(microphone records voice)

(e.g., Fourier transform)



Speech Recognition (2 of 2)

- Frequency spectrogram
  - Basic sounds in the signal (40-50 **phonemes)** (e.g., “a” in “cat”)
    - Template matching against a database of phonemes
      - Using dynamic time warping (speech speed)
  - Constructing words from phonemes (e.g., “th”+”i”+”ng”=thing)
    - Unreliable/probabilistic phonemes (e.g., “th” 50%, “f” 30%, …)
    - Non-unique pronunciations (e.g., tomato),
    - statistics of transitions phonemes/words (hidden Markov models)
- Words

Speech Recognition – Complications

- No simple mapping between sounds and words
  - Variance in pronunciation due to gender, dialect, …
    - Restriction to handle just one speaker
  - Same sound corresponding to diff. words
    - e.g., bear, bare
  - Finding gaps between words
    - "how to recognize speech"
    - "how to wreck a nice beach"
  - Noise

Today

- Full Stages of NLP

1. Speech Recognition
2. Syntax Analysis

3. Semantic Interpretation
    1. Discourse Analysis
4. Pragmatics & World Knowledge

Syntax

Determine how words are put together to form correct sentences

The/DET rose/NOUN is/VERB red/ADJ.

Is/VERB red/ADJ the/DET rose/NOUN.

Express the relations among the constituents of a sentence

Constituents

- also called, syntactic structures or non-terminals

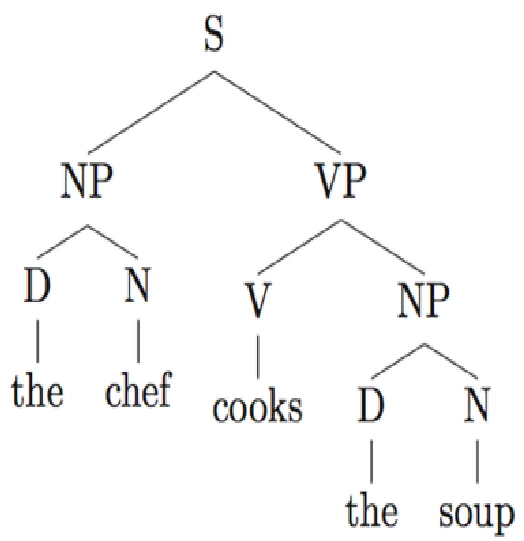
- Main Constituents:
  - S: sentence *The boy is happy.*
  - NP: noun phrase *the little boy from Paris, Sam Smith, I,*
  - VP: verb phrase *eat an apple, sing, leave Paris in the night*
  - PP: prepositional phrase *in the morning, about my ticket*
  - AdjP: adjective phrase *really funny, rather clear*
  - AdvP: adverb phrases *slowly, really slowly*

Syntactic Analysis (1 of 3)

- Rules of syntax (grammar) specify the possible organization of words in sentences and allows us to determine sentence's structure(s)
  - "John saw Mary with a telescope"
    - John saw (Mary with a telescope)
    - John (saw Mary with a telescope)
- Parsing: given a sentence and a grammar
  - Checks that the sentence is correct according with the grammar and if so returns a **parse tree** representing the structure of the sentence

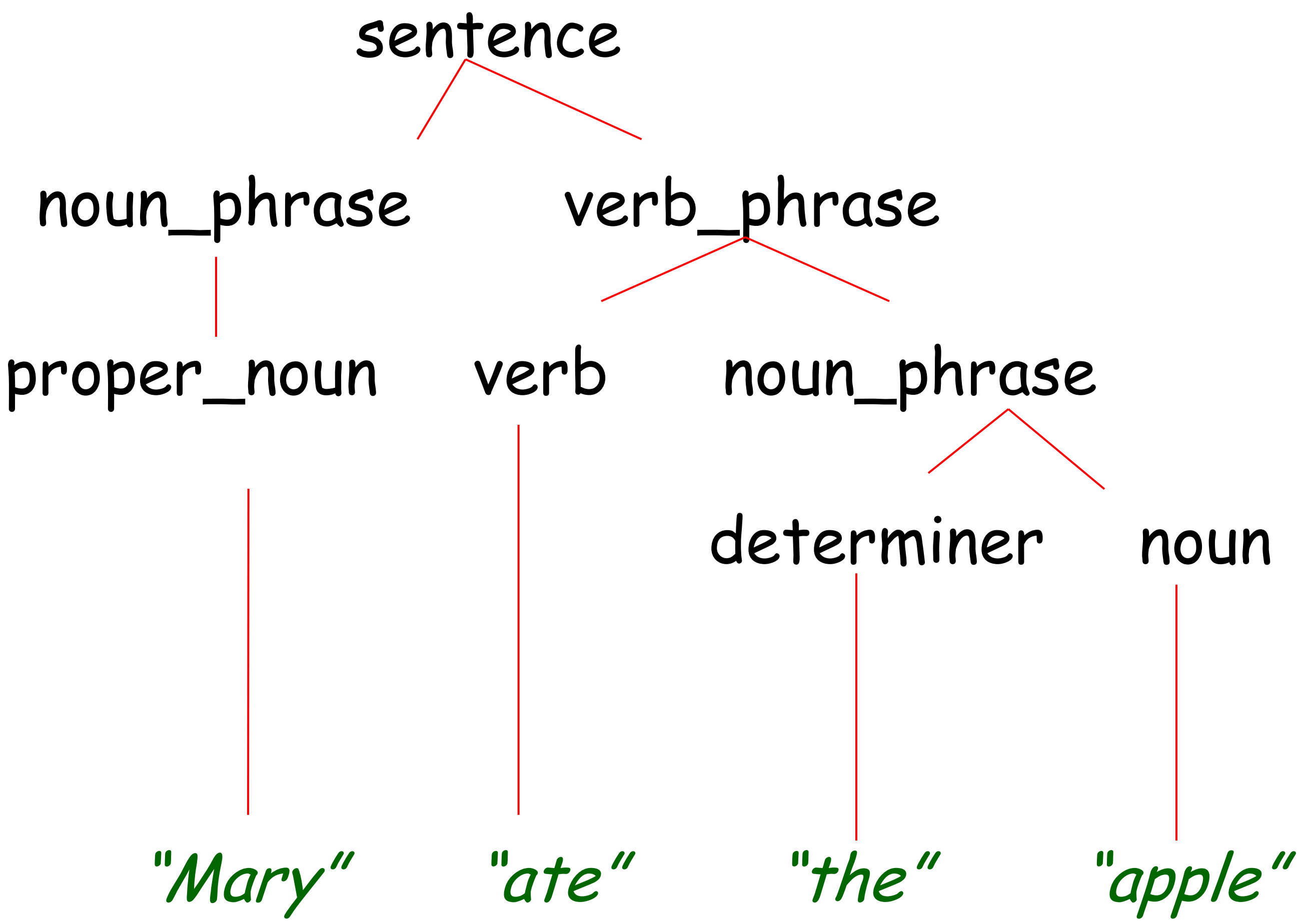
Syntactic Analysis - Grammar (2 of 3)

- sentence -> noun_phrase, verb_phrase
- noun_phrase -> proper_noun
- noun_phrase -> determiner, noun
- verb_phrase -> verb, noun_phrase
- proper_noun -> [mary]
- noun -> [apple]
- verb -> [ate]
- determiner -> [the]



A Parse Tree:
a tree representation of the application of the grammar to a specific sentence.

Syntactic Analysis – Parsing (3 of 3)



Context Free Grammar(CFG)

Consists of:

- set of non-terminal symbols
  - constituents & parts-of-speech
  - S, NP, VP, PP, ART, N, V, ...
- set of terminal symbols
  - lexicon of words & punctuation
  - *cat, mouse, nurses, eat, ...*
- a non-terminal designated as the starting symbol
  - sentence S
- a set of re-write rules
  - having a single non-terminal on the LHS and one or more terminal or non-terminal in the RHS
  - S --> NP VP
  - NP --> Pro | PN | ART N

An Example

- Lexicon:

```
N --> flights | trip | breeze | morning // noun
V --> is | prefer | like                // verb
ADJ --> direct | cheapest | first       // adjective
PRO --> me | I | you | it               // pronoun
PN --> Chicago | United | Los Angeles   // proper noun
ART --> the | a | this                  // determiner
Prep --> from | to | in                 // preposition
```

- Grammar:

```
S --> NP VP                 // I + prefer United
NP --> PRO | PN | ART N     // I | Chicago | the morning
VP --> V | V NP | V NP PP   // is | prefer + United,
PP --> PREP NP              // to Chicago
```

Parsing

- parsing:
  - goal: assign syntactic structures to a sentence
  - result: (set of) parse trees
- we need:
  - a grammar: description of the language constructions
  - a parsing strategy: how the syntactic analysis are to be computed

Parsing Strategies

- parsing is seen as a search problem through the space of all possible parse trees
  - bottom-up (data-directed): words --> grammar
  - top-down (goal-directed): grammar --> words
  - breadth-first: compute all paths in parallel
  - depth-first: exhaust 1 path before considering another

Example: *John ate the cat*

- Bottom-up parsing/ breath first
  - John ate the cat
  - PN ate the cat
  - PN V the cat
  - PN V ART cat
  - PN V ART N
  - NP V ART N
  - NP V NP
  - NP VP
  - S

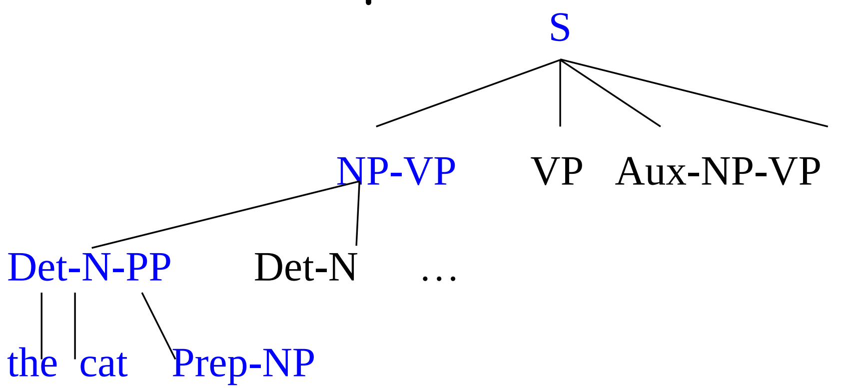
Top-down parsing / depth first

- S
- NP VP
- PN VP
- John VP
- John V NP
- John ate NP
- John ate ART N
- John ate the N
- John ate the cat

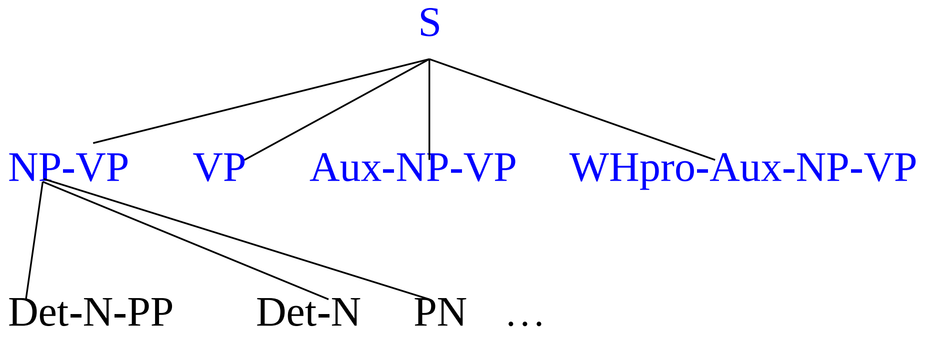
Depth-first vs Breath-first

the cat eats the mouse.

- depth-first: exhaust 1 path before considering another



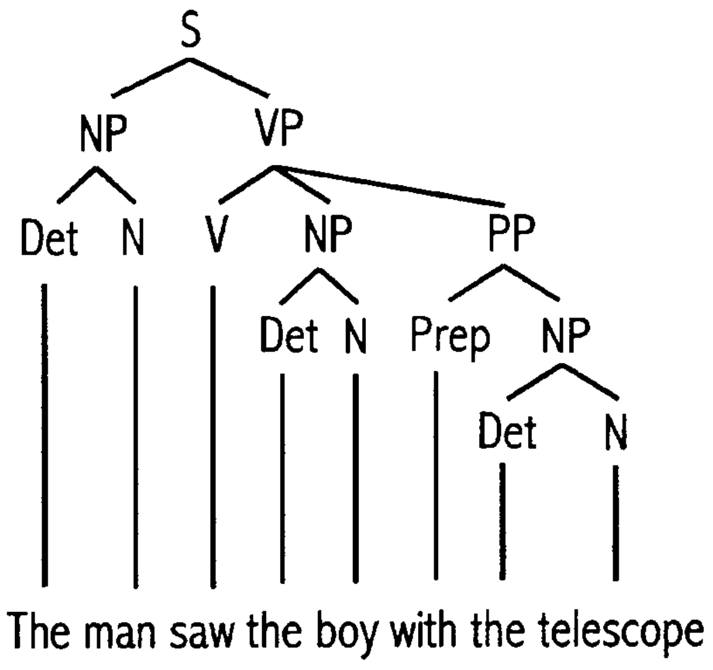
- breadth-first: compute 1 level at a time



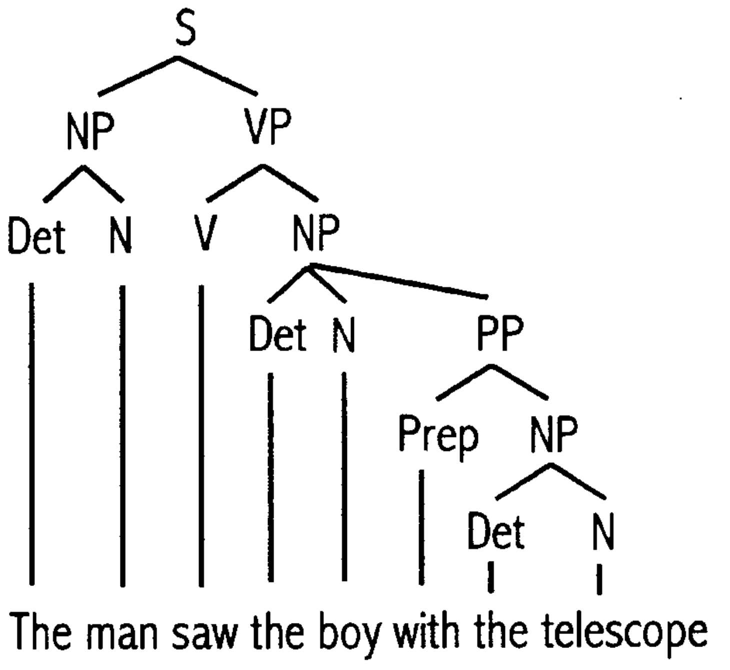
Problem of Redundant Analysis

- Backtracking to try alternative solutions may involve rebuilding structures already built
- Happens very often…
  - garden path sentences
    - *The horse raced past the barn fell*
    - *The prime numbers few.*
    - *Until the police arrest the drug dealers control the street.*
  - prepositional phrase attachment (PP-attachment)
    - *We painted the wall with cracks.*
    - *The man saw the boy with the telescope.*
    - *I shot an elephant in my pyjamas.*

PP attachment:
The man saw the boy with the telescope.



Correct parse 1



Correct parse 2

Difficulties in Parsing

- PP attachment

  The man saw the boy with the telescope.

- NP bracketing

  plastic cat food can cover — *(plastic cat) (food can) cover*
plastic (cat food can) cover
(plastic cat food) (can cover)

- Conjunctions and appositives

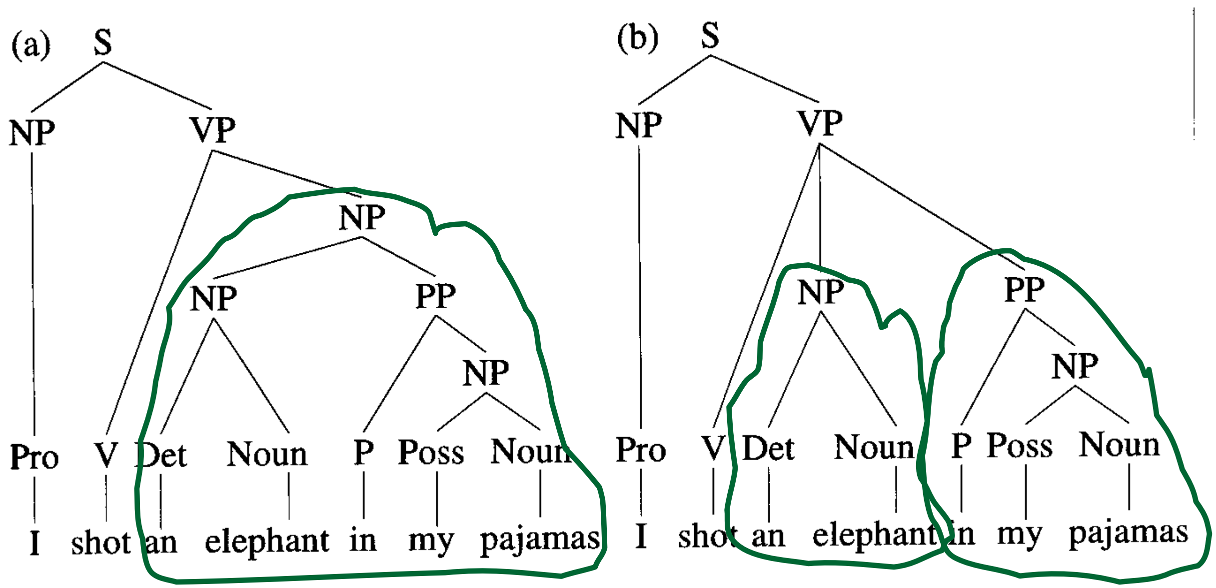
  Maddy, my dog, and Samy — *(Maddy, my dog), and (Samy)*
(Maddy), (my dog), and (Samy)

- These phenomena can quickly increase the number of possible parse trees!

Probabilistic Parsing

"One morning I shot an elephant in my pyjamas. How he got into my pyjamas, I don't know."

G. Marx, *Animal Crackers*, 1930.



Sentences can be very ambiguous…

- A non-probabilistic parser may find a large set of possible parses
- --> need to pick the most probable parse one from the set

Example of a PCFG

$S \rightarrow NP\ VP$	1.0	$NP \rightarrow NP\ PP$	0.4
$PP \rightarrow P\ NP$	1.0	$NP \rightarrow$ *astronomers*	0.1
$VP \rightarrow V\ NP$	0.7	$NP \rightarrow$ *ears*	0.18
$VP \rightarrow VP\ PP$	0.3	$NP \rightarrow$ *saw*	0.04
$P \rightarrow$ *with*	1.0	$NP \rightarrow$ *stars*	0.18
$V \rightarrow$ *saw*	1.0	$NP \rightarrow$ *telescopes*	0.1

Intuitively, P(VP → V NP) is:

- the probability of expanding VP by a V NP, as opposed to any other rules for VP

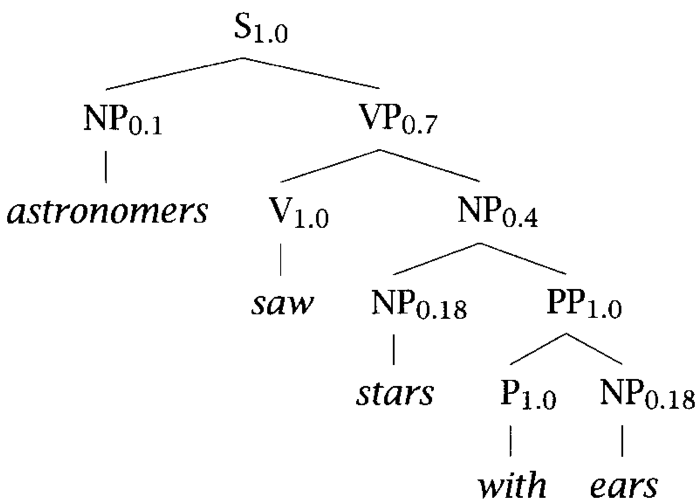
So for:

- VP: $\forall i\ \Sigma_i$ P(VP --> B) = .7 + .3 = 1
- NP: $\forall i\ \Sigma_i$ P(NP --> B) = .4 + .1 + .18 + .04 + .18 + .1 = 1

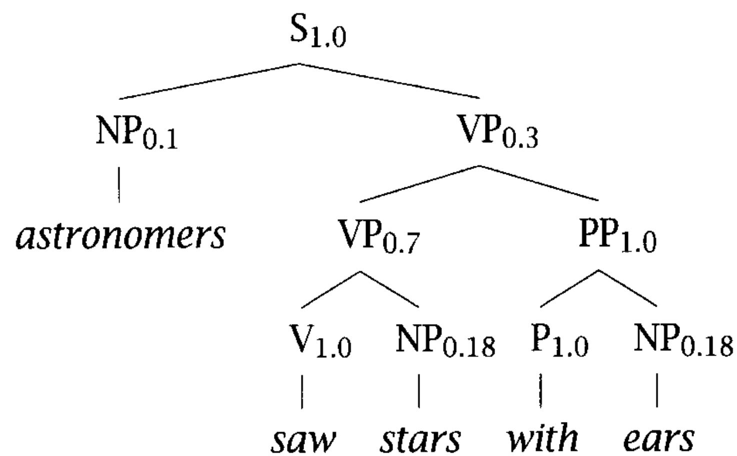
Probability of a parse tree

- Product of the probabilities of the rules used in subtrees
- Ex: *"Astronomers saw stars with ears."*

t_1:



t_2:



$P(t_1) = 1 \times .1 \times .7 \times 1 \times .4 \times .18 \times 1 \times 1 \times .18$
$= .0009072$

$P(t_2) = 1 \times .1 \times .3 \times .7 \times 1 \times 1 \times .18 \times 1 \times .18$
$= .0006804$

Today

- Full Stages of NLP

1. Speech Recognition
2. Syntax Analysis
3. Semantic Interpretation
   1. Discourse Analysis
4. Pragmatics & World Knowledge

Semantic Analysis

- Generates (partial) meaning/ representation of the sentence from its syntactic structure(s).

- Map sentences to some representation of its meaning
  - Representation:
    - Logics, Semantic network, frames,…We'll see this, in the next class.

1. Lexical Semantics:
   - i.e. Meaning of individual words

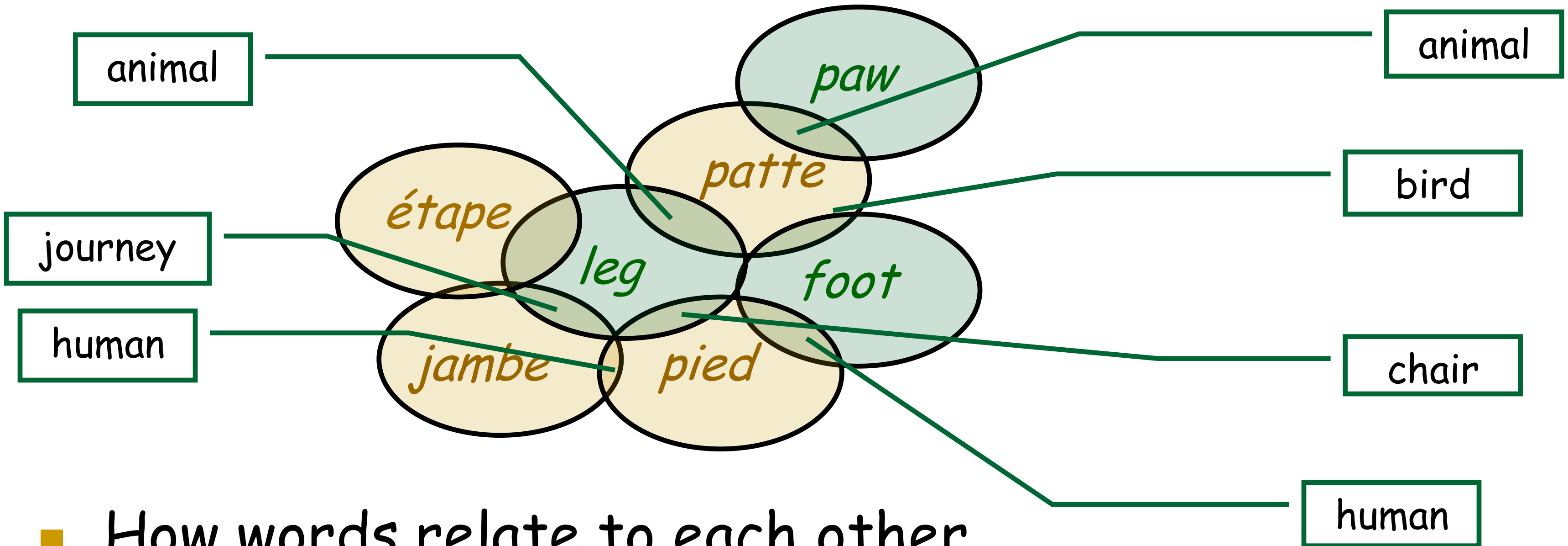
2. Compositional semantics:
   - i.e. meaning of combination of words.

     Sentence: A tall man likes Mary

     Representation: $\exists x$ man(x) & tall(x) & likes(x,mary)

Lexical Semantics

- What a word means (WSD)
  - E.g. Overlapping of word senses across languages



- How words relate to each other
  - E.g. synonymy, antonym, hyponymy, …
    - *How big/large is this plane?*
    - *I think I made a big / ?? large mistake.*

Word Sense Disambiguation (WSD)

- Determining which sense of a word is used in a specific sentence
  - *I went to the bank of Montreal and deposited 50$.*
  - *I went to the bank of the river and dangled my feet.*
- Can be done:
  - with rule-based methods
  - with statistical methods

Rule-based WSD

In our house, everybody has a career and none of them includes washing dishes.
In her tiny kitchen, Ms. Chen works efficiently, stir-frying several simple dishes.

- Use of selectional restrictions: semantic restrictions on the surrounding words
  - predicate *wash* takes as argument an <object>
  - predicate *stir-fry* takes as argument <food>
- BUT:
  - Cannot account for figures of speech:
    - *<animal> drowns in <liquid>*
    - *She is drowning in money*
      metaphor: money is viewed as a liquid
  - Even without figures of speech, it may not be enough:
    - ex: *What kind of dishes do you recommend?*

Probabilistic WSD

- intuition:
  - sense of a word depends on the sense of surrounding words
- ex: *bass* = fish or musical instrument

Surrounding words	**Most probable sense**
...river...	fish
...violin...	instrument
...salmon...	fish
play/Verb + *bass*	instrument
bass + player	instrument
...striped...	fish

Probabilistic WSD

- WSD is viewed as typical classification problem
  - use machine learning techniques (ex. Naïve Bayes classifier) to train a system
  - that learns a classifier (a function *f*) to assign to unseen examples one of a fixed number of senses (categories)

- Input:
  - Target word:
    - The word to be disambiguated
  - Context (feature vector):
    - a vector of relevant linguistic features that represents its context (ex: a window of words around the target word)

What are the features/attributes?

- Take a window of *n* word around the target word
- Encode information about the words around the target word
  - *An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.*

Naïve Bayes Classification

- Goal: choose the most probable sense s^* for a word given a vector V of surrounding words

- vector V contains:
  - Attribute: words around [*fishing, big, sound, player, fly, rod, …*]
  - Value: frequency of words [0, 0, 0, 2, 1, 0, …]

- Bayes decision rule:
  - $s^* = \text{argmax}_{s_k} P(s_k|V)$
  - where:
    - S is the set of possible senses for the target word
    - s_k is a sense in S
    - V is the feature vector (the representation of the context)

Naïve Bayes WSD

$$s^* = \underset{s_k}{\mathrm{argmax}} \left(\log P(s_k) + \sum_{j=1}^{n} \log P(v_j \mid s_k) \right)$$

- Training a Naïve Bayes classifier
  = estimating $P(v_j|s_k)$ and $P(s_k)$ from a sense-tagged training corpus
  = finding the most likely sense k

$$P(v_j \mid s_k) = \frac{\mathrm{count}(v_j, s_k)}{\sum_t \mathrm{count}(v_t, s_k)}$$

Nb of occurrences of feature j over the total nb of features appearing in windows of S_k

$$P(s_k) = \frac{\mathrm{count}(s_k)}{\mathrm{count}(\mathrm{word})}$$

Nb of occurrences of sense k over nb of all occurrences of ambiguous word

Example

- Training corpus (context window = ±3 words):

  ...Today the World **Bank/**BANK1 and partners are calling for greater relief...
  ...Welcome to the **Bank/**BANK1 of America the nation's leading financial institution...
  ...Welcome to America's Job **Bank/**BANK1 Visit our site and...
  ...Web site of the European Central **Bank/**BANK1 located in Frankfurt...
  ...The Asian Development **Bank/**BANK1 ADB a multilateral development finance...

  ...lounging against verdant **banks/**BANK2 carving out the...
  ...for swimming, had warned her off the **banks/**BANK2 of the Potomac. Nobody...

- Training:
  - P(*the*|BANK1) = 5/30 P(*the*|BANK2) = 3/12
  - P(*world*|BANK1) = 1/30 P(*world*|BANK2) = 0/12
  - P(*and*|BANK1) = 1/30 P(*and*|BANK2) = 0/12
  - ...
  - P(*off*|BANK1) = 0/30 P(*off*|BANK2) = 1/12
  - P(*Potomac*|BANK1) = 0/30 P(*Potomac*|BANK2) = 1/12

  - P(BANK1) = 5/7 P(BANK2) = 2/7

- Disambiguation: *"I lost my left shoe on the* ***banks*** *of the river Nile."*

  - Score(BANK1)=log(5/7) + log(P(*shoe*|BANK1))+log(P(*on*|BANK1))+log(P(*the*|BANK1)) ...
  - Score(BANK2)=log(2/7) + log(P(*shoe*|BANK2)+log(P(*on*|BANK2))+log(P(*the*|BANK2)) ...

Compositional Semantics

- *The cat eats the mouse = The mouse is eaten by the cat.*
- Goal:
  - map an expression into a **knowledge representation**
    - a representation of context-independent, literal meaning
    - e.g. first-order predicate logic
  - to assign semantic roles (different from syntactic roles):
    - Agent, Patient, Instrument, Time, Location, …
- E.g.
  - *The child hid the candy under the bed.*

```
Hide ( agent=child, patient=candy,
    location=under_the_bed, time=past)
```

Syntax-Driven Semantic Labeling

Semantic Role	**Definition**	**Example**	**Syntactic Realization**
agent	instigator of the action	*John* broke the window.	subject, *by*-pp
instrument	force/tool causing event	*The hammer* broke the window.	with-pp, subject
patient	thing affected	John broke *the window.*	direct object, subject
experiencer	person involved in perception/state	*John* saw the clouds.	subject
destination	final location	I walked *to school.*	*to*-pp, *into*-pp
…	…	…	…

Some Difficulties

- Syntax is not enough
  - *I ate spaghetti with a fork.* <instrument>
  - *I ate spaghetti with my sister.* <accompanying person>
  - *I ate spaghetti with meat balls.* *<attribute of food>*
  - *I ate spaghetti with lots of appetite.* *<manner>*

  - *Gun* = instrument that can kill
  - *Metal gun…* a gun made out of metal
  - *Water gun…* a gun made out of water?
  - *Fake gun…* it is a gun anyways? Can it kill?

  - *General Kane…* person but *General Motors* … corporation

- Parallel problems to syntactic ambiguity
  - *Happy [cats and dogs] live on the farm*
  - *[Happy cats] and dogs live on the farm*

- Quantifier Scoping
  - *Every man loves a woman.*
    - $\forall$ m ($\exists$ f man(m) $\wedge$ woman(f) $\wedge$ loves(m, f))
    - $\exists$ f($\forall$ m man(m) $\wedge$ woman(f) $\wedge$ loves(m, f))

Today

- Full Stages of NLP

1. Speech Recognition
2. Syntax Analysis
3. Semantic Interpretation
    1. Discourse Analysis

4. Pragmatics & World Knowledge

Discourse Analysis

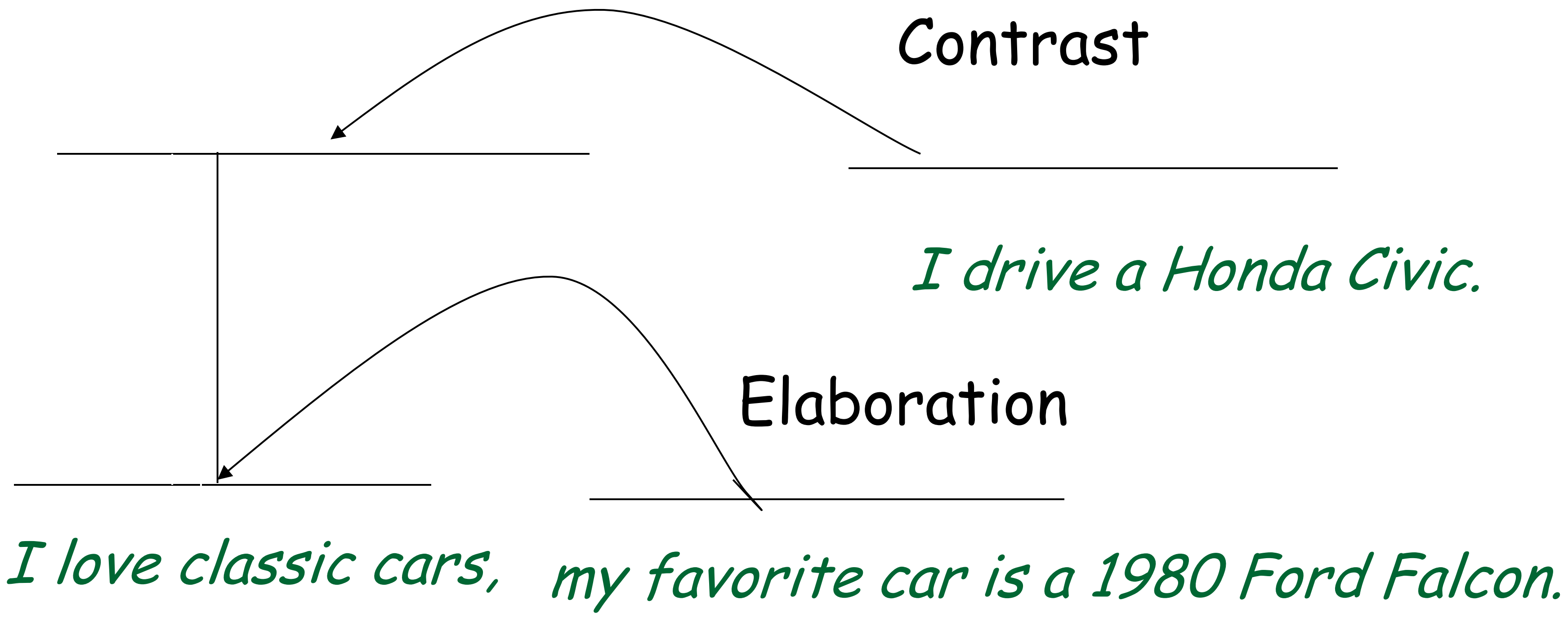
- In logics: $A \wedge B \wedge C \Leftrightarrow C \wedge B \wedge A$
- Not in NL:
  - *John visited Paris. He bought Mary some expensive cologne. Then he flew home. He went to Kmart. He bought some underwear.*
  - *John visited Paris. Then he flew home. He went to Kmart. He bought Mary some expensive cologne. He bought some underwear.*

- Humans infer relations between sentences that may not be explicitly stated in order to make a text coherent.
  - *(*) Bill went to see his mother. The trunk is what makes the bonsai, it gives it both its grace and power.*

Example of Discourse Relations

CONDITION	*If it rains, I will go out.*
SEQUENCE	*Do this, then do that.*
CONTRAST	*This is good, but this is better.*
CAUSE	*Because I was sick, I could not do my assignment.*
RESULT	*Click on the button, the red light will blink.*
PURPOSE	*To use the computer, get an access code.*
ELABORATION	*The solution was developed by Alan Turing. Turing was a great mathematician living in Great Britain. He was an atheist as well as gay.*

A Discourse Tree



Today

- Full Stages of NLP
  1. Speech Recognition
  2. Syntax Analysis
  3. Semantic Interpretation
     1. Discourse Analysis
  4. Pragmatics & World Knowledge

Pragmatics

- Uses context of utterance
  - Where, by who, to whom, why, when it was said
  - Intentions: *inform, request, promise, criticize,* …
- Handling Pronouns
  - "Mary eats apples. She likes them."
    - She="Mary", them="apples".
- Handling ambiguity
  - Pragmatic ambiguity: "you're late": What's the speaker's intention: informing or criticizing?

Pragmatics

- goes beyond the study of the meaning of a sentence
- tries to explain what the speaker is really expressing
- understanding how people use language socially
  - Ex. figures of speech, speech acts, …
  - Ex: *Could you spare some change?*

Using World Knowledge

- Using our general knowledge of the world to interpret a sentence/discourse
- Ex:

  A men was killed yesterday because a jealous husband returned home earlier then usual.

  The professor sent the student to see the principal because…

  …he wanted to see him.

  …he was throwing paper balls in class.

  …he could not take it anymore.

  - *Ex: Silence of the lambs…*

→ Can be solved using scripts or other built-in world-knowledge see: CYC, for example (http://www.opencyc.org/)

we will see this in 2 weeks

Today

- Full Stages of NLP

1. Speech Recognition
2. Syntax Analysis
3. Semantic Interpretation
    1. Discourse Analysis
4. Pragmatics & World Knowledge

References

- Artificial Intelligence: A Modern Approach, By Russell and Norvig, Prentice Hall, 2010.
- Manning, and Schütze, Foundations of Statistical Natural Language Processing, MIT Press (1999)
- Some slides and/or pictures in the following are adapted from slides ©2013. Prof. L. Kosseim;