NLP
Introduction to NLP

NLP Tasks (1/2)
Part of Speech Tagging

The swimmer is getting ready to run in the final race.
The swimmer is getting ready to run in the final race.

- Run – verb or noun?
- Final – noun or adjective?
- Race – verb or noun?
The candidate is preparing for his \textit{run} for the presidency.
The swimmer is getting ready to \textit{run} in the final race.
Parsing

- Myriam slept.
- Myriam wrote a novel.
- Myriam gave Sally flowers.
- Myriam ate pizza with olives.
- Myriam ate pizza with Sally.
- Myriam ate pizza with a fork.
- Myriam ate pizza with remorse.
## Phrase-Structure Grammar

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>S → NP VP</td>
</tr>
<tr>
<td>NP</td>
<td>NP → DET N</td>
</tr>
<tr>
<td>NP</td>
<td>NP → NP PP</td>
</tr>
<tr>
<td>VP</td>
<td>VP → VBD</td>
</tr>
<tr>
<td>VP</td>
<td>VP → VBD NP</td>
</tr>
<tr>
<td>VP</td>
<td>VP → VBD NP NP</td>
</tr>
<tr>
<td>PP</td>
<td>PP → PRP NP</td>
</tr>
<tr>
<td>DET</td>
<td>DET → the</td>
</tr>
<tr>
<td>DET</td>
<td>DET → that</td>
</tr>
<tr>
<td>DET</td>
<td>DET → a</td>
</tr>
<tr>
<td>N</td>
<td>N → child</td>
</tr>
<tr>
<td>N</td>
<td>N → window</td>
</tr>
<tr>
<td>N</td>
<td>N → car</td>
</tr>
<tr>
<td>VBD</td>
<td>VBD → found</td>
</tr>
<tr>
<td>VBD</td>
<td>VBD → ate</td>
</tr>
<tr>
<td>VBD</td>
<td>VBD → saw</td>
</tr>
<tr>
<td>PRP</td>
<td>PRP → in</td>
</tr>
<tr>
<td>PRP</td>
<td>PRP → of</td>
</tr>
<tr>
<td>PRP</td>
<td>PRP → through</td>
</tr>
</tbody>
</table>
The child saw the car through the window.
Stanford Parser

Please enter a sentence to be parsed:

Housing starts, the number of new homes being built, rose 7.2% in March to an annual rate of 549,000 units, up from a revised 512,000 in February, the Commerce Department said.

Language: English  Sample Sentence  Perse

Your query

Housing starts, the number of new homes being built, rose 7.2% in March to an annual rate of 549,000 units, up from a revised 512,000 in February, the Commerce Department said.

Tagging

Housing/NN starts/NNS ,/, the/DT number/NN of/IN new/JJ homes/NNS being/VBG built/VBN ,/, rose/VBD 7.2/CD 4/NNS in/IN March/NNP to/TD an/DT annual/JJ rate/NN of/IN 549,000/CD units/NNS ,/, up/PP from/IN a/DT revised/VBD 512,000/CD in/IN February/NNP ,/, the/DT Commerce/NNS Department/NNS said/VBD ./.
Housing starts, the number of new homes being built, rose 7.2% in March to an annual rate of 549,000 units, up from a revised rate of 512,000 in February, the Commerce Department said.
This Problem is Pretty // Easy

• Commercial for a phone company
• Garden path sentences
  – Don’t bother coming
  – Don’t bother coming early
  – Take the turkey out of the oven at five
  – Take the turkey out of the oven at five to four
  – I got canned
  – I got canned peaches for dinner
  – All Americans need to buy a house
  – All Americans need to buy a house is a lot of money

• Can you think of more such examples?
Solution

• This problem is pretty // easy

• Criteria
  – The part before // should be a complete sentence
  – The full sentence has a different meaning than the part before //
  – The part before // should not already be ambiguous
Dependency Parsing

likes

Mary apples

yellow
IL-2 and IL-15 induced the production of IL-17 and IFN-\(\gamma\) by PBMCs in a dose dependent manner.
Housing starts rose 7.2% in March, up 549,000 units from a revised 512,000 in February, said the Department of Commerce.

Rate of annual home building rose to 512,000 units from 512,000 in February.
• RESEARCH ALERT—Wells Fargo cuts PPD Inc to market perform
• China Southern Air Upgraded To Overweight From Neutral—HSBC
• CITIGROUP RAISES INGERSOLL RAND <IR.N> TO HOLD FROM SELL
• TCF Financial Corp Raised To Overweight From Neutral By JPMorgan
• BAIRD CUTS KIOR INC <KIOR.O> TO UNDERPERFORM RATING
• BRIEF—RESEARCH ALERT—Global Equities Research cuts LinkedIn to equal weight
# Information Extraction

<table>
<thead>
<tr>
<th>DATE/TIME</th>
<th>TICKER</th>
<th>COMPANY</th>
<th>SOURCE</th>
<th>OLD</th>
<th>NEW</th>
<th>CHANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PPD Inc</td>
<td>Wells Fargo</td>
<td></td>
<td>market perform</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>China Southern Air</td>
<td>HSBC</td>
<td>Neutra l</td>
<td>Overweight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR.N</td>
<td>INGERSOLL RAND</td>
<td>CITIGROUP</td>
<td>SELL</td>
<td>HOLD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TCF Financial Corp</td>
<td>JPMorgan</td>
<td>Neutra l</td>
<td>Overweight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIOR.O</td>
<td>KIOR INC</td>
<td>BAIRD</td>
<td></td>
<td>UNDERPERFORM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LinkedIn</td>
<td></td>
<td>Global Equities Research</td>
<td></td>
<td>equal weight</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
False Positives

• Examples of false positives
  – BARCLAYS CUTS FLAGSTONE REINSURANCE <FSR.N> PRICE TARGET TO $9 FROM $11
  – Rimage To Buy Qumu For $52M;; Raises Dividend;; Lowers EPS View
  – S&P rates Ameren Illinois commercial paper 'A–3'
  – BRIEF–Moody's changes otlk for Stirling Water Seafield Finance to positive
  – BRIEF–RESEARCH ALERT–HSBC cuts price targets on European telcos
  – Stifel cuts Philip Morris price target
  – Media General shares plummet on Moody's downgrade

• Explain why these are false positives.
Answers to the Quiz

- **BARCLAYS CUTS FLAGSTONE REINSURANCE <FSR.N> PRICE TARGET TO $9 FROM $11**
  - Didn’t cut the ratings but the price target

- **Rimage To Buy Qumu For $52M;; Raises Dividend;; Lowers EPS View**
  - Lowers eps view

- **S&P rates Ameren Illinois commercial paper 'A-3‘**
  - Debt rating

- **BRIEF–Moody's changes otlk for Stirling Water Seafield Finance to positive**
  - Changes outlook

- **BRIEF–RESEARCH ALERT–HSBC cuts price targets on European telcos**
  - Not a company but a group of companies

- **Stifel cuts Philip Morris price target**
  - Price target, not rating

- **Media General shares plummet on Moody's downgrade**
  - Event in the past
Semantics

- First order logic
- Inference
- Semantic analysis

\( \forall x, y: \text{Mother} (x, y) \implies \text{Parent} (x, y) \)
NACLO Problem

• “Bertrand and Russell”, 2014 problem by Ben King
  – to be covered in a later lecture
NACLO Solution

• Bertrand and Russell
Reading Comprehension

Mars Polar Lander - Where Are You?

(January 18, 2000) After more than a month of searching for a signal from NASA’s Mars Polar Lander, mission controllers have lost hope of finding it. The Mars Polar Lander was on a mission to Mars to study its atmosphere and search for water, something that could help scientists determine whether life ever existed on Mars. Polar Lander was to have touched down December 3 for a 90-day mission. It was to land near Mars’ south pole. The lander was last heard from minutes before beginning its descent. The last effort to communicate with the three-legged lander ended with frustration at 8 a.m. Monday. “We didn’t see anything,” said Richard Cook, the spacecraft’s project manager at NASA’s Jet Propulsion Laboratory. The failed mission to the Red Planet cost the American government more than $200 million dollars. Now, space agency scientists and engineers will try to find out what could have gone wrong. They do not want to make the same mistakes in the next mission.


1. When did the mission controllers lose hope of communicating with the lander?
   Answer: 8AM, Monday Jan. 17, 2000
2. Who is the Polar Lander’s project manager?
3. Where on Mars was the spacecraft supposed to touch down?
4. What did the Mars Global Surveyor do?
5. What was the mission of the Mars Polar Lander?
There are four bungalows in our cul-de-sac. They are made from these materials: straw, wood, brick and glass.

Mrs. Scott's bungalow is somewhere to the left of the wooden one and the third one along is brick. Mrs. Umbrella owns a straw bungalow and Mr. Tinsley does not live at either end, but lives somewhere to the right of the glass bungalow. Mr. Wilshaw lives in the fourth bungalow, whilst the first bungalow is not made from straw.

Who lives where, and what is their bungalow made from?

Word Sense Disambiguation

• “The thieves took off with 100 gold bars”.  
  – Did they steal 100 drinking establishments?  
  – Or 100 measures of a song?
Word Sense Disambiguation

Bar=Noun
S: (n) barroom, bar, saloon, ginmill, taproom (a room or establishment where alcoholic drinks are served over a counter) "he drowned his sorrows in whiskey at the bar"
S: (n) bar (a counter where you can obtain food or drink) "he bought a hot dog and a coke at the bar"
S: (n) bar (a rigid piece of metal or wood; usually used as a fastening or obstruction or weapon) "there were bars in the windows to prevent escape"
S: (n) measure, bar (musical notation for a repeating pattern of musical beats) "the orchestra omitted the last twelve bars of the song"
S: (n) bar (an obstruction (usually metal) placed at the top of a goal) "it was an excellent kick but the ball hit the bar"
S: (n) prevention, bar (the act of preventing) "there was no bar against leaving"; "money was allocated to study the cause and prevention of influenza"
S: (n) bar ((meteorology) a unit of pressure equal to a million dynes per square centimeter) "unfortunately some writers have used bar for one dyne per square centimeter"
S: (n) bar (a submerged (or partly submerged) ridge in a river or along a shore) "the boat ran aground on a submerged bar in the river"
S: (n) legal profession, bar, legal community (the body of individuals qualified to practice law in a particular jurisdiction) "he was admitted to the bar in New Jersey"
S: (n) stripe, streak, bar (a narrow marking of a different color or texture from the background) "a green toad with small black stripes or bars"; "may the Stars and Stripes forever wave"
S: (n) cake, bar (a block of solid substance (such as soap or wax)) "a bar of chocolate"
S: (n) Browning automatic rifle, BAR (a portable .30 caliber automatic rifle operated by gas pressure and fed by cartridges from a magazine; used by United States troops in World War I and in World War II and in the Korean War)
S: (n) bar (a horizontal rod that serves as a support for gymnasts as they perform exercises)
S: (n) bar (a heating element in an electric fire) "an electric fire with three bars"
S: (n) bar ((law) a railing that encloses the part of the courtroom where the judges and lawyers sit and the case is tried) "spectators were not allowed past the bar"

Bar=Verb
S: (v) bar, debar, exclude (prevent from entering; keep out) "He was barred from membership in the club"
S: (v) barricade, block, blockade, stop, block off, block up, bar (render unsuitable for passage) "block the way"; "barricade the streets"; "stop the busy
WSD is Important for Translation

• Paul plays soccer
  – Paul joue au football

• Paul plays the guitar
  – Paul joue de la guitare

• “wall” in German
  – die Chinesische Mauer (The Great Wall of China)
  – (otherwise Wand)

• “wall” in Spanish
  – pared, muro, muralla
Wolff, currently a journalist in Argentina, played with Del Bosque in the final years of the seventies in Real Madrid.

- http://cogcomp.cs.illinois.edu/page/demo_view/NER
- http://nlp.stanford.edu:8080/ner/

The distribution of myeloid lineage-associated cytokine receptors and lysosomal proteins was analyzed in human CD34+ cord blood cell (CB) subsets at different stages of myeloid commitment by reverse-transcriptase polymerase chain reaction (RT-PCR). The highly specific granulomonocyte-associated lysosomal proteins myeloperoxidase (MPO) and lysozyme (LZ), as well as the transcription factor PU.1, were already detectable in the most immature CD34+Thy-1+ subset. Messenger RNA (mRNA) levels for the granulocyte-colony stimulating factor (G-CSF)
NLP