Introduction to NLP

Semantic Role Labeling
Syntactic Variation

- Last week, Min broke the window with a hammer.
- The window was broken with a hammer by Min last week.
- With a hammer, Min broke the window last week.
- Last week, the window was broken by Min with a hammer.
- Min broke the window.
- The window broke.
- The window was broken with a hammer.
Semantic Role Labeling

• Determining
  – who
  – did what
  – to whom
  – when
  – where
  – why
  – how

• Uses
  – Question answering
  – Machine translation
  – Text summarization
Case Theory (Fillmore 1968)

• **Agent**
  – Actor of an action
  – The musician performed a new piece

• **Patient**
  – Entity affected by the action
  – Samantha hurt her hand

• **Instrument**
  – Tool used in performing action
  – Min broke the window with a hammer

• **Beneficiary**
  – Entity for whom action is performed
  – The mother bought ice cream for the children

• **Source**
  – Origin of the affected entity
  – I got the book from my friend

• **Destination**
  – Destination of the affected entity
Using syntactic information

• Syntactic information
  – “by X” for agent
  – “with X” for instrument

• Exceptions
  – “by car”
  – “with pleasure”
SRL task

• Input
  The teacher gave the test to the students in the morning.

• Output
  \([\text{The teacher}]_{\text{AGENT}} \text{ gave } [\text{the test}]_{\text{OBJ}} \text{ to } [\text{the students}]_{\text{RECIP}} \text{ in the morning}]_{\text{TMP}}.\)
Illinois Demo

Text Analysis Demo

Input Text:
Chartered as the Illinois Industrial University, the University opened with its current name in 1868 as one of the original 37 public land-grant institutions created after Abraham Lincoln signed the Morrill Act in 1862.

Output:

http://cogcomp.cs.illinois.edu/page/demo_view/SRL
<table>
<thead>
<tr>
<th>Word</th>
<th>PoS</th>
<th>Base Chunk</th>
<th>Clause</th>
<th>Syntactic Tree</th>
<th>Named Entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>and</td>
<td>CC</td>
<td>*</td>
<td>(S*)</td>
<td>*</td>
<td>(AM-DIS*)</td>
</tr>
<tr>
<td>to</td>
<td>TO</td>
<td>(VP*)</td>
<td>(S*)</td>
<td>(S(VP*))</td>
<td>(AM-DIS*)</td>
</tr>
<tr>
<td>attract</td>
<td>VB</td>
<td>*</td>
<td>(VP*)</td>
<td>*</td>
<td>(AM-PNC*)</td>
</tr>
<tr>
<td>younger</td>
<td>JJ</td>
<td>(NP*)</td>
<td>*</td>
<td>(NP*)</td>
<td>*</td>
</tr>
<tr>
<td>listeners</td>
<td>NNS</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Radio</td>
<td>NNP</td>
<td>(NP*)</td>
<td>(NP*)</td>
<td>(ORG*)</td>
<td>(A0*)</td>
</tr>
<tr>
<td>Free</td>
<td>NNP</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Europe</td>
<td>NNP</td>
<td>*</td>
<td>*</td>
<td>(A1*)</td>
<td>*</td>
</tr>
<tr>
<td>intersperses</td>
<td>VBZ</td>
<td>(VP*)</td>
<td>(VP*)</td>
<td>intersperse</td>
<td>(V*)</td>
</tr>
<tr>
<td>the</td>
<td>DT</td>
<td>(NP*)</td>
<td>(NP(NP*))</td>
<td>*</td>
<td>(A1*)</td>
</tr>
<tr>
<td>latest</td>
<td>JJS</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>in</td>
<td>IN</td>
<td>(PP*)</td>
<td>(PP*)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Western</td>
<td>JJ</td>
<td>(NP*)</td>
<td>(NP*)</td>
<td>(MISC*)</td>
<td>*</td>
</tr>
<tr>
<td>rock</td>
<td>NN</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>groups</td>
<td>NNS</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Figure 1: An example of an annotated sentence, in columns. Input consists of words (1st column), PoS tags (2nd), base chunks (3rd), clauses (4th), full syntactic tree (5th) and named entities (6th). The 7th column marks target verbs, and their propositions are found in remaining columns. According to the PropBank Frames, for attract (8th), the A0 annotates the attractor, and the A1 the thing attracted; for intersperse (9th), A0 is the arranger, and A1 the entity interspersed.

[Carreras and Marquez 2005]
FrameNet

- Berkeley
- Chuck Fillmore
- https://framenet.icsi.berkeley.edu/
PropBank

- PropBank
  - U. Colorado
  - Martha Palmer
  - http://verbs.colorado.edu/~mpalmer/projects/ace.html
  - Arg0 usually agent
  - Arg1 usually patient/theme
  - 13 labels for Adjuncts (Time, Location, Manner)
PropBank Example

- **Roleset id: break.01**, break, cause to not be whole
- **Roles:**
  - Arg0: breaker (vnrole: 23.2–agent, 40.8.3–1–experiencer, 45.1–agent)
  - Arg1: thing broken (vnrole: 23.2–patient1, 40.8.3–1–patient, 45.1–patient)
  - Arg2: instrument (vnrole: 45.1–instrument)
  - Arg3: pieces (vnrole: 23.2–patient2)
- **Example: just transitive**
- Stock prices rallied as the Georgia-Pacific bid broke the market's recent gloom.
  - Arg0: the Georgia-Pacific bid
  - Rel: broke
  - Arg1: the market's recent gloom
PropBank Example

• **Example: with instrument**
  John broke the window with a rock.
  Arg0: John
  Rel: broke
  Arg1: the window
  Arg2: with a rock

• **Example: with pieces**
  John broke the window into a million pieces.
  Arg0: John
  Rel: broke
  Arg1: the window
  Arg3: into a million pieces

• **Example: inchoative**
  The window broke into a million pieces.
  Arg1: The window
  Rel: broke
  Arg3: into a million pieces
Papers

- Gildea and Jurafsky 2002
- Xue and Palmer 2004
- Punakyanok et al. 2004
- Pradhan et al. 2004
- Yi and Palmer 2005
- Marquez et al. 2005
- Haghighi et al. 2005
Approaches

• Selectional restrictions
  – instruments should be tools (e.g., *not* “with pleasure”)
  – agents and beneficiaries should be animate (e.g., not “for a reason”)

• Use WordNet
  – “the teacher” is a person is animate

• Parse node classification
Features Used (1)

- Phrase type
- Governing category
- Parse tree path (e.g., $N \uparrow NP \uparrow S \downarrow VP \downarrow V$)
- Position (e.g., does the phrase precede or follow the predicate)
- Voice
- Head word
- Subcategorization
- Argument Set
- Argument Order

List from Palmer, Gildea, and Xue 2010
Features Used (2)

- Previous Role
- Head Word Part of Speech
- Named Entities in Constituents
- Verb Clustering
- Head Words of Objects of PPs
- First/Last Word/POS in Constituent
- Constituent Order
- Constituent Tree Distance
- Temporal Cue Words

List from Palmer, Gildea, and Xue 2010
Results

• CONLL Shared task (since 2004)
• Best performance over 80% F1 measure
NLP