Lecture 18
Syntactic Dependencies

Intro to NLP, CS585, Fall 2014
http://people.cs.umass.edu/~brenocon/inlp2014/
Brendan O’Connor
Error analysis: election forecasting

Senate Polls Overstate Democratic Support Across the Board

In almost every race, the polling average skewed Democratic, sometimes by a substantial margin.

Competitive evaluation

How the Senate Forecast Models Did

As measured by the logarithmic score, one commonly used technique for scoring probabilistic forecasts, on Election Day most of the forecasts performed roughly the same, with the models from the Washington Post and Daily Kos scoring the highest. For much of the year, the Upshot’s forecast scored at or near the top of the pack.

Logarithmic Score

Dependencies on their own

http://universaldependencies.github.io/docs/#language-en

Parsing to dependencies

- One approach: parse to constituents, then convert.
  - Appears to be most accurate method, for English

- Alternative: direct dependency parsing
  - Advantages: training data availability, algorithms sometimes simpler, no need for converter
  - Disadvantages: may lose deeper syntax information encoded in constituency tree

- Methods for dependency parsing
  - Discriminative approaches are most popular
  - Graph-based: predict whole tree.
  - Transition-based (shift-reduce): incrementally predict left-to-right.
Graph-based parsing

Inference: dynamic programming, minimum spanning trees...
Learning: structured perceptron (or similar)
Graph-based parsing

Current research: how to use higher order features
Decoding is more difficult

Inference: dynamic programming, minimum spanning trees...
Learning: structured perceptron (or similar)
# Arc-Eager Transition System [Nivre 2003]

Configuration: \((S, B, A)\) \([S = \text{Stack}, B = \text{Buffer}, A = \text{Arcs}]\)

Initial: \([(\[]), [0, 1, \ldots, n], \{ \} )\]

Terminal: \((S, [ ], A)\)

Shift: \((S, i|B, A) \Rightarrow (S|i, B, A)\)

Reduce: \((S|i, B, A) \Rightarrow (S, B, A)\)

Right-Arc\( (k)\): \((S|i, j|B, A) \Rightarrow (S|i|j, B, A \cup \{(i, j, k)\})\)

Left-Arc\( (k)\): \((S|i, j|B, A) \Rightarrow (S, j|B, A \cup \{(j, i, k)\})\) \(\neg h(i, A) \land i \neq 0\)

Notation:

- \(S|i = \text{stack with top } i \text{ and remainder } S\)
- \(j|B = \text{buffer with head } j \text{ and remainder } B\)
- \(h(i, A) = i \text{ has a head in } A\)
Example Transition Sequence

[ROOT] \_5 \ [Economic, news, had, little, effect, on, financial, markets, .] \_B

ROOT Economic news had little effect on financial markets .

adj noun verb adj noun prep adj noun .
Example Transition Sequence

\[
\begin{align*}
&\text{ROOT, Economic}_S \quad \text{news, had, little, effect, on, financial, markets, .}_B \\
&\text{ROOT} \quad \text{Economic} \quad \text{news} \quad \text{had} \quad \text{little} \quad \text{effect} \quad \text{on} \quad \text{financial} \quad \text{markets} \quad . \\
&\text{adj} \quad \text{noun} \quad \text{verb} \quad \text{adj} \quad \text{noun} \quad \text{prep} \quad \text{adj} \quad \text{noun} \quad .
\end{align*}
\]
Example Transition Sequence

\[(\text{ROOT}_S \ [\text{news, had, little, effect, on, financial, markets, .}]_B)\]

ROOT Economic news had little effect on financial markets .

amod

adj noun verb adj noun prep adj noun .

Recent Advances in Dependency Parsing
Example Transition Sequence

[ROOT, news]ₜ [had, little, effect, on, financial, markets, .]ₜ

ROOT Economic news had little effect on financial markets .

adj noun verb adj noun prep adj noun .
Example Transition Sequence

\[ \text{ROOT}_S \quad [\text{had, little, effect, on, financial, markets, .}]_B \]

```
ROOT Economic news had little effect on financial markets .
  adj  noun  verb  adj  noun   prep  adj  noun
  amod nsubj
```
Transition-Based Dependency Parsing

Example Transition Sequence

\[ [\text{ROOT}, \text{had}]_S \] \[ [\text{little, effect, on, financial, markets, .}]_B \]

```
root

ROOT  Economic  news  had  little  effect  on  financial  markets  .
adj  noun  verb  adj  noun  prep  adj  noun  .
```
Example Transition Sequence

\[ \text{ROOT, had, little}_S \quad \text{effect, on, financial, markets, .}_B \]

- root
- amod
- nsubj

ROOT  Economic  news  had  little  effect  on  financial  markets  .

adj  noun  verb  adj  noun  prep  adj  noun  .
Example Transition Sequence

\[ [\text{ROOT, had}]_S \quad [\text{effect, on, financial, markets, .}]_B \]

- **root**
- **amod**
- **nsubj**
- **amod**

**ROOT**: Economic
**adj**: news
**noun**: had
**verb**: little
**adj**: effect
**noun**: on
**prep**: financial
**adj**: markets
**noun**: .
Example Transition Sequence

\[ \text{ROOT, had, effect}_S \quad \text{on, financial, markets, .}_B \]

\[
\text{ROOT} \quad \text{Economic} \quad \text{news} \quad \text{had} \quad \text{little} \quad \text{effect} \quad \text{on} \quad \text{financial} \quad \text{markets} \quad .
\]
Example Transition Sequence

\[ \text{ROOT, had, effect, on, financial}_S \quad \text{[markets, .]}_B \]
Example Transition Sequence

\[ \text{ROOT, had, effect, on}_S \quad \text{[markets, ]}_B \]

```
root

ROOT          Economic   news   had   little   effect   on   financial   markets .
            adj         noun   verb   adj      noun   prep   adj      noun.
  amod      nsubj     amod     prep      amod
```

Transitions:  
1. \( S \rightarrow \text{ROOT, had, effect, on} \)  
2. \( \text{ROOT, had, effect, on} \rightarrow \text{markets, } \)
Example Transition Sequence

\[ [\text{ROOT, had, effect, on, markets}]_S \rightarrow [.]_B \]

\[
\begin{align*}
\text{ROOT} & \quad \text{Economic} \\
\text{adj} & \quad \text{news} \\
\text{noun} & \quad \text{had} \\
\text{verb} & \quad \text{little} \\
\text{adj} & \quad \text{effect} \\
\text{noun} & \quad \text{on} \\
\text{prep} & \quad \text{financial} \\
\text{adj} & \quad \text{markets} \\
\text{noun} & \quad .
\end{align*}
\]
Example Transition Sequence

[ROOT, had, effect, on]_s [.]_B

ROOT  Economic
  adj   noun
  had   verb
  little
  effect
  on
  financial
  markets
  .

Recent Advances in Dependency Parsing

[Slides: McDonald and Nivre, EACL 2014 tutorial]

Thursday, November 6, 14
Example Transition Sequence

\[ \text{ROOT, had, effect} \]_S \quad [.]_B

```
ROOT
   amod
   nsubj
ROOT  Economic
adj  news  noun  had
verb  little
adj  effect  noun
  prep  on
  pmod  financial
adj  markets
  amod
```
Example Transition Sequence

\[ \text{ROOT, had}]_S \quad [.]_B \]

- **ROOT**
- **amod**
- **nsubj**
- **dobj**
- **prepp**
- **pmod**

- Economic adj
- news noun
- had verb
- little adj
- effect noun
- on prep
- financial adj
- markets noun
- .
Example Transition Sequence

\[
\text{EXAMPLE} \quad \text{SEQUENCE}
\]

\[
\text{ROOT, had, .}_S \quad \text{[]}_B
\]
Greedy Inference

- Given an oracle $o$ that correctly predicts the next transition $o(c)$, parsing is deterministic:

  $\text{Parse}(w_1, \ldots, w_n)$
  \begin{align*}
  1 & \quad c \leftarrow ([ ]_S, [0, 1, \ldots, n]_B, \{ \}) \\
  2 & \quad \textbf{while } B_c \neq [] \\
  3 & \quad t \leftarrow o(c) \\
  4 & \quad c \leftarrow t(c) \\
  5 & \quad \textbf{return } G = ([0, 1, \ldots, n], A_c)
  \end{align*}

- Complexity given by upper bound on number of transitions
- Parsing in $O(n)$ time for the arc-eager transition system
Feature Representation

- Features over input tokens relative to $S$ and $B$

**Configuration**

**Features**
- $\text{word}(S_2) = \text{ROOT}$
- $\text{word}(S_1) = \text{had}$
- $\text{word}(S_0) = \text{effect}$
- $\text{word}(B_0) = \text{on}$
- $\text{word}(B_1) = \text{financial}$
- $\text{word}(B_2) = \text{markets}$
Feature Representation

- Features over input tokens relative to $S$ and $B$
- Features over the (partial) dependency graph defined by $A$

Configuration

Features

\[
\begin{align*}
\text{dep}(S_1) &= \text{root} \\
\text{dep}(\text{lc}(S_1)) &= \text{nsubj} \\
\text{dep}(\text{rc}(S_1)) &= \text{dobj} \\
\text{dep}(S_0) &= \text{dobj} \\
\text{dep}(\text{lc}(S_0)) &= \text{amod} \\
\text{dep}(\text{rc}(S_0)) &= \text{NIL}
\end{align*}
\]
Feature Representation

- Features over input tokens relative to $S$ and $B$
- Features over the (partial) dependency graph defined by $A$
- Features over the (partial) transition sequence

**Configuration**

```
[ROOT, had, effect]_S  [on, financial, markets, .]_B

ROOT Economic news had little effect on financial markets .
```

**Features**

- $t_{i-1} = \text{Right-Arc(dobj)}$
- $t_{i-2} = \text{Left-Arc(amod)}$
- $t_{i-3} = \text{Shift}$
- $t_{i-4} = \text{Right-Arc(root)}$
- $t_{i-5} = \text{Left-Arc(nsubj)}$
- $t_{i-6} = \text{Shift}$
Feature Representation

- Features over input tokens relative to \( S \) and \( B \)
- Features over the (partial) dependency graph defined by \( A \)
- Features over the (partial) transition sequence

Configuration

\[
\text{ROOT, had, effect}_S \quad \text{on, financial, markets}_B
\]

Features

\[
\begin{align*}
 t_{i-1} &= \text{Right-Arc(dobj)} \\
 t_{i-2} &= \text{Left-Arc(amod)} \\
 t_{i-3} &= \text{Shift} \\
 t_{i-4} &= \text{Right-Arc(root)} \\
 t_{i-5} &= \text{Left-Arc(nsubj)} \\
 t_{i-6} &= \text{Shift}
\end{align*}
\]

- Feature representation unconstrained by parsing algorithm
Local Learning

- Given a treebank:
  - Reconstruct oracle transition sequence for each sentence
  - Construct training data set \( D = \{(c, t) \mid o(c) = t\} \)
  - Maximize accuracy of local predictions \( o(c) = t \)
- Any (unstructured) classifier will do (SVMs are popular)
- Training is local and restricted to oracle configurations
Greedy, Local, Transition-Based Parsing

- **Advantages:**
  - Highly efficient parsing – linear time complexity with constant time oracles and transitions
  - Rich history-based feature representations – no rigid constraints from inference algorithm

- **Drawback:**
  - Sensitive to search errors and error propagation due to greedy inference and local learning

- The major question in transition-based parsing has been how to improve learning and inference, while maintaining high efficiency and rich feature models
Beam Search

- Maintain the $k$ best hypotheses [Johansson and Nugues 2006]:

  $$\text{Parse}(w_1, \ldots, w_n)$$

  1. $$\text{Beam} \leftarrow \{(S, [0, 1, \ldots, n]_B, \{ \})\}$$
  2. $$\text{while } \exists c \in \text{Beam} \ [B_c \neq \{\}]$$
  3. $$\text{foreach } c \in \text{Beam}$$
  4. $$\text{foreach } t$$
  5. $$\text{Add}(t(c), \text{NewBeam})$$
  6. $$\text{Beam} \leftarrow \text{Top}(k, \text{NewBeam})$$
  7. $$\text{return } G = (\{0, 1, \ldots, n\}, A_{\text{Top}(1, \text{Beam})})$$

- **Note:**
  - $$\text{Score}(c_0, \ldots, c_m) = \sum_{i=1}^{m} w \cdot f(c_{i-1}, t_i)$$
  - Simple combination of locally normalized classifier scores
  - Marginal gains in accuracy
An oracle can be approximated by a (linear) classifier:

\[ o(c) = \arg\max_t w \cdot f(c, t) \]

- History-based feature representation \( f(c, t) \)
- Weight vector \( w \) learned from treebank data
State of the art

- Unlabeled attachment scores as of 2014
- Accuracy of choose-the-parent
- Labeled scores a little lower
- Datasets vary in quality, so take with a grain of salt

<table>
<thead>
<tr>
<th>Language</th>
<th>Best Published</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic</td>
<td>81.12 (MS11)</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>94.02 (ZH13)</td>
</tr>
<tr>
<td>Chinese</td>
<td>92.68 (LX14)</td>
</tr>
<tr>
<td>Czech</td>
<td>91.04 (ZL14)</td>
</tr>
<tr>
<td>Danish</td>
<td>92.00 (ZH13)</td>
</tr>
<tr>
<td>Dutch</td>
<td>86.47 (ZL14)</td>
</tr>
<tr>
<td>English</td>
<td>93.22 (MA13)</td>
</tr>
<tr>
<td>German</td>
<td>92.41 (MA13)</td>
</tr>
<tr>
<td>Japanese</td>
<td>93.74 (LX14)</td>
</tr>
<tr>
<td>Portuguese</td>
<td>93.03 (KR10)</td>
</tr>
<tr>
<td>Slovene</td>
<td>86.95 (MS11)</td>
</tr>
<tr>
<td>Spanish</td>
<td>88.24 (ZL14)</td>
</tr>
<tr>
<td>Swedish</td>
<td>91.62 (ZH13)</td>
</tr>
<tr>
<td>Turkish</td>
<td>77.55 (KR10)</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>89.58</strong></td>
</tr>
</tbody>
</table>

• Transition-based dependency parsers: extremely fast (e.g. MaltParser)

• Does syntax encode interesting semantics?
I want to go to New York on Sunday
I want to go to New York on Sunday
Natural Language Understanding

Shallow Semantics: *Frames and Roles*

I want to go to New York on Sunday

Encodes an event or scenario

[Slides: Dipanjan Das]
Natural Language Understanding

Shallow Semantics: *Frames and Roles*

I want to go to New York on Sunday

[Slides: Dipanjan Das]

Thursday, November 6, 14
Natural Language Understanding

Shallow Semantics: *Frames and Roles*

participant or role for the frame

Traveler

I want to go to New York on Sunday

[Slides: Dipanjan Das]
I want to go to New York on Sunday.
I want to go to New York on Sunday.
Semantics: MRs

• For question-answering, dialogue systems, story understanding, etc… one subproblem: want a relational meaning representation

• (Why relational?)

• Predicate-Argument structures
  • e.g. \( V(S, O) \): verb has noun arguments
  • (\(~\text{Verb}) \) Actions/Events/Frames, \( hating \)
  • (\(~\text{Noun}) \) Roles/Slots/Arguments
# Example

<table>
<thead>
<tr>
<th>Text</th>
<th>I saw a person</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SVO syntactic structures</strong></td>
<td><code>see(I, person)</code></td>
</tr>
<tr>
<td></td>
<td><code>[verb=see, subj=I, directobj=person]</code></td>
</tr>
<tr>
<td><strong>Semantic roles</strong></td>
<td><code>[event=see, agent=I, patient=person]</code></td>
</tr>
</tbody>
</table>
**Example**

<table>
<thead>
<tr>
<th>Text</th>
<th>I saw a person</th>
</tr>
</thead>
</table>
| Feature-structure (frame-style?) representation | type= SeeingEvent  
  time= Past  
  subj= [word= I,  
          grampers=1st,  
          num= sg]  
  ... |

(High-level syntax like LFG / HPSG?)  
(Or is it low-level semantics?)
Example

<table>
<thead>
<tr>
<th>Text</th>
<th>I believe I saw a person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame-style representation</td>
<td></td>
</tr>
</tbody>
</table>

TopCtx =>

- event = believe
- agent = I
- theme = BeliefCtx

BeliefCtx =>

- event = see
- agent = I
- patient = person

ctx(TopCtx)
ctx(BeliefCtx)
inctx(TopCtx, event(believe))
inctx(TopCtx, agent(believe, I))
inctx(TopCtx, theme(believe, BeliefCtx))
inctx(BeliefCtx, event(see))
inctx(BeliefCtx, agent(see, I))
inctx(BeliefCtx, patient(see, person))

(Factivity via Davidsonian semantics, description/modal logic formalism: Bobrow et al 2005)