Semantic Role Labeling
CS 224N, Spring 2009
Christopher Manning

Slides mainly from a tutorial from Scott Wen-tau Yih and Kristina Toutanova (Microsoft Research), with additional slides from Sameer Pradhan (BBN) as well as Dan Jurafsky and myself.
Syntactic Variations versus Semantic Roles

Yesterday, Kristina hit Scott with a baseball
Scott was hit by Kristina yesterday with a baseball
Yesterday, Scott was hit with a baseball by Kristina
With a baseball, Kristina hit Scott yesterday
Yesterday Scott was hit by Kristina with a baseball
The baseball with which Kristina hit Scott yesterday was hard
Kristina hit Scott with a baseball yesterday

Agent, hitter
Patient, Thing hit
Instrument
Temporal adjunct
Syntactic Variations (as trees)
Semantic Role Labeling – Giving Semantic Labels to Phrases

- \([AGENT \ John] \ broke \ [THEME \ the \ window]\)
- \([THEME \ The \ window] \ broke\)
- \([AGENT \ Sotheby’s] \ .. \ offered \ [RECIPIENT \ the \ Dorrance \ heirs] \ [THEME \ a \ money-back \ guarantee]\)
- \([AGENT \ Sotheby’s] \ offered \ [THEME \ a \ money-back \ guarantee] \ to \ [RECIPIENT \ the \ Dorrance \ heirs]\)
- \([THEME \ a \ money-back \ guarantee] \ offered \ by \ [AGENT \ Sotheby’s]\)
- \([RECIPIENT \ the \ Dorrance \ heirs] \ will \ [ARM-NEG \ not] \ be \ offered \ [THEME \ a \ money-back \ guarantee]\)
Some typical semantic roles

<table>
<thead>
<tr>
<th>Thematic Role</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGENT</td>
<td>The volitional causer of an event</td>
</tr>
<tr>
<td>EXPERIENCER</td>
<td>The experiencer of an event</td>
</tr>
<tr>
<td>FORCE</td>
<td>The non-volitional causer of the event</td>
</tr>
<tr>
<td>THEME</td>
<td>The participant most directly affected by an event</td>
</tr>
<tr>
<td>RESULT</td>
<td>The end product of an event</td>
</tr>
<tr>
<td>CONTENT</td>
<td>The proposition or content of a propositional event</td>
</tr>
<tr>
<td>INSTRUMENT</td>
<td>An instrument used in an event</td>
</tr>
<tr>
<td>BENEFICIARY</td>
<td>The beneficiary of an event</td>
</tr>
<tr>
<td>SOURCE</td>
<td>The origin of the object of a transfer event</td>
</tr>
<tr>
<td>GOAL</td>
<td>The destination of an object of a transfer event</td>
</tr>
</tbody>
</table>
Some typical semantic roles

<table>
<thead>
<tr>
<th>Thematic Role</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGENT</td>
<td><em>The waiter</em> spilled the soup.</td>
</tr>
<tr>
<td>EXPERIENCER</td>
<td><em>John</em> has a headache.</td>
</tr>
<tr>
<td>FORCE</td>
<td><em>The wind</em> blows debris from the mall into our yards.</td>
</tr>
<tr>
<td>THEME</td>
<td>Only after Benjamin Franklin broke <em>the ice</em>...</td>
</tr>
<tr>
<td>RESULT</td>
<td>The French government has built a <em>regulation-size baseball diamond</em>...</td>
</tr>
<tr>
<td>CONTENT</td>
<td>Mona asked “<em>You met Mary Ann at a supermarket?</em>”</td>
</tr>
<tr>
<td>INSTRUMENT</td>
<td>He turned to poaching catfish, stunning them <em>with a shocking device</em>...</td>
</tr>
<tr>
<td>BENEFICIARY</td>
<td>Whenever Ann Callahan makes hotel reservations <em>for her boss</em>...</td>
</tr>
<tr>
<td>SOURCE</td>
<td>I flew in <em>from Boston</em>.</td>
</tr>
<tr>
<td>GOAL</td>
<td>I drove <em>to Portland</em>.</td>
</tr>
</tbody>
</table>
What is SRL good for?

**Question Answering**

Q: What was the name of the first computer system that defeated Kasparov?
A: [PATIENT Kasparov] was defeated by [AGENT Deep Blue] [TIME in 1997].

Q: When was Napoleon defeated?
Look for: [PATIENT Napoleon] [PRED defeat-synset] [ARGM-TMP *ANS*]

More generally:
What is SRL good for?
Applications as a simple meaning rep’n

- Machine Translation
  - English (SVO)
    - [AGENT The little boy]
    - [PRED kicked]
    - [THEME the red ball]
    - [ARGM-MNR hard]
  - Farsi (SOV)
    - [AGENT pesar koocholo]  boy-little
    - [THEME toop germezi]   ball-red
    - [ARGM-MNR moqtam]      hard-adverb
    - [PRED zaad-e]           hit-past

- Document Summarization
  - Predicates and Heads of Roles summarize content

- Information Extraction
  - SRL can be used to construct useful rules for IE
Application: Semantically precise search

Query: afghans destroying opium poppies

Results 1 - 10 of about 829 for afghans destroying opium poppies. (0.07 seconds)

Japan Today - News - Afghans threaten to grow more opium poppies ...
Afghans threaten to grow more opium poppies. ... 30 JST ISLAMABAD — Growers of opium poppies in Afghanistan's ... cultivation if compensation for destroying the most ...
www.japantoday.com/gidx/news228842.html - 10k - Cached - Similar pages

MSNBC - Afghans on losing side of the drug war
... Afghans on losing side of the drug war. ... makes a lot more sense to grow poppies and opium instead of ... The government has a modest goal of destroying 30 percent of ...
msnbc.msn.com/id/4991545/ - 41k - Cached - Similar pages

NewsHour Extra: Afghans Vote in First Democratic Election ...
... About 7 million Afghans now farm poppy for economic ... role in the manufacture and sale of opium. ... only criminalizing the Afghan economy. destroying our agriculture ...
www.pbs.org/newshour/extra/features/july-dec04/afghanistan_10-25_printout.html - 9k - Cached - Similar pages

(PDF) Letters from Afghanistan
File Format: PDF/Adobe Acrobat - View as HTML
... This weapon is the opium poppy, used to produce heroin ... American embassy who fear that the Afghans "are in ... warn that it is no good destroying opium unless there ...
tdh.ch/cms/fileadmin/site_uploads/d/pdf/ projekte/asien/Afghanistan/Letters_from_Afghanistan_17.pdf - Similar pages

Afgha.com - Afghan narcotics add to woes
... 9Provinces that never grew poppies are growing ... in the American embassy who fear that the Afghans "are in ... warn that it is no good destroying opium unless there ...
www.afgha.com/?af=printnews&sdi=40568 - 7k - Cached - Similar pages

NewsCentralAsia - Drugs in Afghanistan: Of carts and horses
Diathesis alternations

John  broke the window.
AGENT  THEME

John  broke the window with a rock.
AGENT  THEME  INSTRUMENT

The rock  broke the window.
INSTRUMENT  THEME

The window broke.
THEME

The window was broken by John.
THEME  AGENT

Doris  gave the book to Cary.
AGENT  THEME  GOAL

Doris  gave Cary the book.
AGENT  GOAL  THEME
Problems with semantic roles

- It’s very hard to produce a formal definition of a role
- There are all sorts of arbitrary role splits
- Intermediary instruments (1-2) vs. enabling instruments (3-4):
  1. The cook opened the jar with the new gadget
  2. The new gadget opened the jar
  3. Sally ate the sliced banana with a fork
  4. *The fork ate the sliced banana
Solutions to the difficulty of defining semantic roles

- Ignore semantic role labels, and just mark arguments of individual verbs as 0, 1, 2
  - PropBank
- Define semantic role labels for a particular semantic domain
  - FrameNet
PropBank

- A corpus of labeled sentences (Penn Treebank WSJ)
- The arguments of each verb are labeled with numbers rather than names, though there are verb frame files:

(19.29) **agree.01**

Arg0: Agreer
Arg1: Proposition
Arg2: Other entity agreeing
Ex1: [Arg0 The group] *agreed* [Arg1 it wouldn’t make an offer unless it had Georgia Gulf’s consent].
Ex2: [ArgM-TMP Usually] [Arg0 John] *agrees* [Arg2 with Mary] [Arg1 on everything].

(19.30) **fall.01**

Arg1: Logical subject, patient, thing falling
Arg2: Extent, amount fallen
Arg3: start point
Arg4: end point, end state of arg1
Ex1: [Arg1 Sales] *fell* [Arg4 to $251.2 million] [Arg3 from $278.7 million].
Ex2: [Arg1 The average junk bond] *fell* [Arg2 by 4.2%].
Proposition Bank (PropBank) 
Frame Files

- **hit.01 “strike”**
  - A0: agent, hitter; A1: thing hit;
    A2: instrument, thing hit by or with
  - \([A_0 \text{Kristina}] \text{hit}[A_1 \text{Scott}] [A_2 \text{with a baseball}] \text{yesterday.}\)

- **look.02 “seeming”**
  - A0: seemer; A1: seemed like; A2: seemed to
  - \([A_0 \text{It}] \text{looked}[A_2 \text{to her}] \text{like}[A_1 \text{he deserved this}]\).

- **deserve.01 “deserve”**
  - A0: deserving entity; A1: thing deserved;
    A2: in-exchange-for
  - \(\text{It looked to her like}[A_0 \text{he}] \text{deserved}[A_1 \text{this}]\).
Proposition Bank (PropBank)
Add a Semantic Layer

Kristina hit Scott with a baseball yesterday.
“The worst thing about him,” said Kristina, “is his laziness.”

[\text{A}_1 \text{ The worst thing about him}] \text{ said } [\text{A}_0 \text{ Kristina }] [\text{C-A}_1 \text{ is his laziness}].
Proposition Bank (PropBank) Final Notes

- **Current release** (Mar 4, 2005): Proposition Bank I
  - Verb Lexicon: 3,324 frame files
  - Annotation: ~113,000 propositions
    [http://www.cis.upenn.edu/~mpalmer/project_pages/ACE.htm](http://www.cis.upenn.edu/~mpalmer/project_pages/ACE.htm)

- **Alternative format**: CoNLL-04,05 shared task
  - Represented in table format
  - Has been used as standard data set for the shared tasks on semantic role labeling
Other Corpora

- **Chinese PropBank**  http://www.cis.upenn.edu/~chinese/cpb/
  - Similar to PropBank, it adds a semantic layer onto Chinese Treebank

- **NomBank**  http://nlp.cs.nyu.edu/meyers/NomBank.html
  - Label arguments that co-occur with nouns in PropBank
  - `[A0 Her] [REL gift] of [A1 a book] [A2 to John]`
CoNLL format

1. lie("he",...)  
2. leak("he", "information obtained from ... he supervised")  
3. obtain(X, "information", "from a wiretap he supervised")  
4. supervise("he", "a wiretap")

<table>
<thead>
<tr>
<th>He</th>
<th>-</th>
<th>(AO*)</th>
<th>(AO*)</th>
<th>*</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>is</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>also</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>accused</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>of</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>lying</td>
<td>lie</td>
<td>(V*)</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>under</td>
<td>-</td>
<td>(AM-LOC*)</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>oath</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>and</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>of</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>leaking</td>
<td>leak</td>
<td>*</td>
<td>(V*)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>information</td>
<td>-</td>
<td>*</td>
<td>(A1*)</td>
<td>(A1*)</td>
<td>*</td>
</tr>
<tr>
<td>obtained</td>
<td>obtain</td>
<td>*</td>
<td>*</td>
<td>(V*)</td>
<td>*</td>
</tr>
<tr>
<td>from</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>a</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>(A2*)</td>
<td>(A1*)</td>
</tr>
<tr>
<td>wiretap</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>he</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>(AO*)</td>
<td></td>
</tr>
<tr>
<td>supervised</td>
<td>supervise</td>
<td>*</td>
<td>*)</td>
<td>*)</td>
<td>(V*)</td>
</tr>
<tr>
<td>.</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
FrameNet [Fillmore et al. 01]

Frame: Hit_target (hit, pick off, shoot)

- **Agent**: Kristina
- **Target**: Scott
- **Instrument**: with a baseball
- **Time**: yesterday

**Lexical units (LUs):** Words that evoke the frame (usually verbs)

**Frame elements (FEs):** The involved semantic roles

http://framenet.icsi.berkeley.edu
FrameNet

- A frame is a semantic structure based on a set of participants and events
- Consider the “change_position_on_scale” frame

<table>
<thead>
<tr>
<th>VERBS:</th>
<th>dwindle</th>
<th>advance</th>
<th>edge</th>
<th>rise</th>
<th>diminish</th>
<th>growth</th>
<th>increase</th>
<th>jump</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>move</td>
<td>mushroom</td>
<td>plummet</td>
<td>reach</td>
<td>fluctuate</td>
<td>rocket</td>
<td>shift</td>
<td>slide</td>
</tr>
<tr>
<td></td>
<td>soar</td>
<td>swell</td>
<td>swing</td>
<td>triple</td>
<td>tumble</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>escalation</td>
<td>explosion</td>
<td>fall</td>
<td>fluctuation</td>
<td>g</td>
<td>growth</td>
<td>hike</td>
<td></td>
</tr>
<tr>
<td></td>
<td>shift</td>
<td>tumble</td>
<td></td>
<td></td>
<td>increase</td>
<td>rise</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ADVERBS:</th>
<th>increasingly</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOUNS:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Roles in this frame

<table>
<thead>
<tr>
<th>Core Roles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ATTRIBUTE</strong></td>
</tr>
<tr>
<td><strong>DIFFERENCE</strong></td>
</tr>
<tr>
<td><strong>FINAL_STATE</strong></td>
</tr>
<tr>
<td><strong>FINAL_VALUE</strong></td>
</tr>
<tr>
<td><strong>INITIAL_STATE</strong></td>
</tr>
<tr>
<td><strong>INITIAL_VALUE</strong></td>
</tr>
<tr>
<td><strong>ITEM</strong></td>
</tr>
<tr>
<td><strong>VALUE_RANGE</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Some Non-Core Roles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DURATION</strong></td>
</tr>
<tr>
<td><strong>SPEED</strong></td>
</tr>
<tr>
<td><strong>GROUP</strong></td>
</tr>
</tbody>
</table>
Examples / Quiz question

- [ITEM Oil] rose [ATTRIBUTE in price] [DIFFERENCE by 2%].
- [ITEM It] has increased [FINAL_STATE to having them 1 day a month].
- [ITEM Microsoft shares] fell [FINAL_VALUE to 7 5/8].
- [Colon cancer incidence] fell [by 50%] [among men].
- a steady increase [from 9.5] [to 14.3] [in dividends]
- a [5%] [dividend] increase

Give the roles for the 3 items with blanks

- Core: Attribute, Difference, Final_State, Final_Value, Initial_State, Initial_Value, Item, Value_Range
  - An Item has a scalar Attribute which moves in the Value_Range portion of a scale
- Some Non-Core: Duration, Speed, Group
  - The Item changes its Attribute within a Group for a certain Duration or at a Speed
Problems with FrameNet

- Example sentences are chosen by hand
  - Not randomly selected
  - Complete sentences not labeled
- Since TreeBank wasn’t used
  - No perfect parses for each sentence
- Still ongoing (that’s good and bad)
Some History

- Fillmore 1968: The case for case
  - Proposed semantic roles as a shallow semantic representation
- Simmons 1973:
  - Built first automatic semantic role labeler
    - Based on first parsing the sentence
FrameNet vs PropBank

FRAMENET ANNOTATION:
[Buyer Chuck] bought [Goods a car] [Seller from Jerry] [Payment for $1000].
[Seller Jerry] sold [Goods a car] [Buyer to Chuck] [Payment for $1000].

PROP BANK ANNOTATION:
[Arg0 Chuck] bought [Arg1 a car] [Arg2 from Jerry] [Arg3 for $1000].
[Arg0 Jerry] sold [Arg1 a car] [Arg2 to Chuck] [Arg3 for $1000].
FrameNet vs PropBank -2

FRAMENET ANNOTATION:

\[ \text{Goods A car} \text{ was } bought \ [\text{Buyer by Chuck}] \].

\[ \text{Goods A car} \text{ was } sold \ [\text{Buyer to Chuck} \ [\text{Seller by Jerry}] \].

\[ \text{Buyer Chuck} \text{ was } sold \ [\text{Goods a car} \ [\text{Seller by Jerry}] \].

PROPBANK ANNOTATION:

\[ \text{Arg1 A car} \text{ was } bought \ [\text{Arg0 by Chuck}] \].

\[ \text{Arg1 A car} \text{ was } sold \ [\text{Arg2 to Chuck} \ [\text{Arg0 by Jerry}] \].

\[ \text{Arg2 Chuck} \text{ was } sold \ [\text{Arg1 a car} \ [\text{Arg0 by Jerry}] \].
# Information Extraction versus Semantic Role Labeling

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>IE</th>
<th>SRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>narrow</td>
<td>broad</td>
</tr>
<tr>
<td>Depth of semantics</td>
<td>shallow</td>
<td>shallow</td>
</tr>
<tr>
<td>Directly connected to application</td>
<td>sometimes</td>
<td>no</td>
</tr>
</tbody>
</table>
Overview of SRL Systems

- Definition of the SRL task
  - Evaluation measures
- General system architectures
- Machine learning models
  - Features & models
  - Performance gains from different techniques
Subtasks

- **Identification:**
  - Very hard task: to separate the argument substrings from the rest in this exponentially sized set
  - Usually only 1 to 9 (avg. 2.7) substrings have labels ARG and the rest have NONE for a predicate

- **Classification:**
  - Given the set of substrings that have an ARG label, decide the exact semantic label

- **Core argument** semantic role labeling: (easier)
  - Label phrases with core argument labels only. The modifier arguments are assumed to have label NONE.
Evaluation Measures

Correct: \([A_0 \text{ The queen}] \text{ broke} [A_1 \text{ the window}] \ [\text{AM-TMP yesterday}]\)

Guess: \([A_0 \text{ The queen}] \text{ broke the} [A_1 \text{ window}] \ [\text{AM-LOC yesterday}]\)

<table>
<thead>
<tr>
<th>Correct</th>
<th>Guess</th>
</tr>
</thead>
<tbody>
<tr>
<td>{The queen} \rightarrow A0</td>
<td>{The queen} \rightarrow A0</td>
</tr>
<tr>
<td>{the window} \rightarrow A1</td>
<td>{window} \rightarrow A1</td>
</tr>
<tr>
<td>{yesterday} \rightarrow AM-TMP</td>
<td>{yesterday} \rightarrow AM-LOC</td>
</tr>
<tr>
<td>all other \rightarrow NONE</td>
<td>all other \rightarrow NONE</td>
</tr>
</tbody>
</table>

- Precision, Recall, F-Measure \(\{tp=1, fp=2, fn=2\}\) \(p=r=f=1/3\)
- Measures for subtasks
  - Identification (Precision, Recall, F-measure) \(\{tp=2, fp=1, fn=1\}\) \(p=r=f=2/3\)
  - Classification (Accuracy) acc = .5 (labeling of correctly identified phrases)
  - Core arguments (Precision, Recall, F-measure) \(\{tp=1, fp=1, fn=1\}\) \(p=r=f=1/2\)
What’s the problem with these evaluations?

- Approximating human evaluations is dangerous
  - Humans don’t always agree
  - Not clear if it’s good for anything
  - Sometimes called the “match-a-linguist” task
- What’s a better evaluation?
Basic Architecture of a Generic SRL System

Sentence s, predicate p

annotations
(adding features)

s, p, A

local scoring

s, p, A
score(l|c,s,p,A)

Joint scores take into account dependencies among the labels of multiple phrases

Local scores for phrase labels do not depend on labels of other phrases

semantic roles

joint scoring
SRL architecture: Walk the tree, labeling each parse tree node

- Given a parse tree $t$, label the nodes (phrases) in the tree with semantic labels

Alternative approach: labeling chunked sentences..

[\text{[NP Yesterday]} , \text{[NP Kristina]} \text{[VP hit]} \text{[NP Scott]} \text{[PP with]} \text{[NP a baseball]}].
Why this parse-tree architecture?

Semantic role chunks tend to correspond to syntactic constituents

Propbank:
- 96% of arguments = 1 (gold) parse tree constituent
- 90% of arguments = 1 (Charniak) parse tree constituent
- Simple rules can recover missing 4-10%

FrameNet,
- 87% of arguments = 1 (Collins) parse tree constituent

Why?
- they were labeled from parse trees
- by humans trained in syntax
Parsing Algorithm

- Use a syntactic parser to parse the sentence
- For each predicate (non-copula verb)
  - For each node in the syntax tree
    - Extract a feature vector relative to the predicate
    - Classify the node
  - Do second-pass informed by global info
Baseline Features [Gildea & Jurafsky, 2000]

- Predicate (verb)
- Path from constituent to predicate
- Phrase type (syntactic)
- Position (before/after)
- Voice (active/passive)
- Head Word
- Sub-categorization
Pradhan et al. (2004) Features

- Predicate cluster
- Noun head and POS of PP constituent
- Verb sense
- Partial path
- Named entities in constituent (7) [Surdeanu et al., 2003]
- Head word POS [Surdeanu et al., 2003]
- First and last word in constituent and their POS
- Parent and sibling features
- Constituent tree distance
- Ordinal constituent position
- Temporal cue words in constituent
- Previous 2 classifications
Predicate cluster, automatic or WordNet

He talked for about 20 minutes.

Predicates: spoke, lectured, chatted, explained
Noun Head and POS of PP

He talked for about 20 minutes

Sameer Pradhan
Named Entities and Head Word POS

[Surdeanu et al., 2003]
First and Last Word and POS
Parent and Sibling features

[Diagram showing syntactic relationships between words, with labels for parent and left sibling features.]
Constituent tree distance

He talked for about 20 minutes
Ordinal constituent position

First NP to the right of the predicate

He talked for about 20 minutes
**Temporal Cue Words (~50)**

<table>
<thead>
<tr>
<th>term</th>
<th>term</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>years;ago</td>
</tr>
<tr>
<td>recently</td>
<td>night</td>
</tr>
<tr>
<td>days</td>
<td>hour</td>
</tr>
<tr>
<td>end</td>
<td>decade</td>
</tr>
<tr>
<td>period</td>
<td>late</td>
</tr>
</tbody>
</table>
Previous 2 classifications
Combining Identification and Classification Models

**Step 1. Pruning.**
Using a hand-specified filter.

**Step 2. Identification.**
Identification model (filters out candidates with high probability of NONE)

**Step 3. Classification.**
Classification model assigns one of the argument labels to selected nodes (or sometimes possibly NONE)
Combining Identification and Classification Models – Continued

\[ P(l|c, t, p) = P_{ID}(Id(l) | \Phi(c, t, p)) \times P_{CLS}(l | Id(l), \Phi(c, t, p)) \]

or

\[ P(l|c, t, p) = P(l | \Phi(c, t, p)) \]

One Step.
Simultaneously identify and classify using \( P(l|c, t, p) \)
Joint Scoring Models

- These models have scores for a whole labeling of a tree (not just individual labels)
  - Encode some dependencies among the labels of different nodes

\[ P_{JOINT}(l_1, \ldots, l_n | n, t, p)! = \prod_i P(l_i | n_i, t, p) \]
Combining Local and Joint Scoring Models

- **Tight integration of local and joint scoring in a single probabilistic model** and exact search [Cohn&Blunsom 05] [Màrquez et al. 05], [Thompson et al. 03]
  - When the joint model makes strong independence assumptions

- **Re-ranking** or approximate search to find the labeling which maximizes a combination of local and a joint score [Gildea&Jurafsky 02] [Pradhan et al. 04] [Toutanova et al. 05]
  - Usually exponential search required to find the exact maximizer

- **Exact search for best assignment by local model satisfying hard joint constraints**
  - Using Integer Linear Programming [Punyakanok et al 04,05] (worst case NP-hard)
Joint Scoring: Enforcing Hard Constraints

- **Constraint 1**: Argument phrases do not overlap
  
  *By [A1 working [A1 hard], he] said, you can achieve a lot.*
  
  - Pradhan et al. (04) – greedy search for a best set of non-overlapping arguments
  - Toutanova et al. (05) – exact search for the best set of non-overlapping arguments (dynamic programming, linear in the size of the tree)
  - Punyakanok et al. (05) – exact search for best non-overlapping arguments using integer linear programming

- **Other constraints** ([Punyakanok et al. 04, 05])
  
  - no repeated core arguments (good heuristic)
  - phrases do not overlap the predicate
Joint Scoring: Integrating Soft Preferences

- There are many statistical tendencies for the sequence of roles and their syntactic realizations
  - When both are before the verb, AM-TMP is usually before A0
  - Usually, there aren’t multiple temporal modifiers
  - Many others which can be learned automatically
Joint Scoring: Integrating Soft Preferences

- Gildea and Jurafsky (02) – a smoothed relative frequency estimate of the probability of frame element multi-sets:
  \[ P(\{A0, AM_{TMP}, A1, AM_{TMP}\}|hit) \]
  - Gains relative to local model 59.2 \rightarrow 62.9 FrameNet automatic parses

- Pradhan et al. (04) – a language model on argument label sequences (with the predicate included)
  \[ P( A0, AM_{TMP}, hit, A1, AM_{TMP} ) \]
  - Small gains relative to local model for a baseline system 88.0 \rightarrow 88.9 on core arguments PropBank correct parses

- Toutanova et al. (05) – a joint model based on CRFs with a rich set of joint features of the sequence of labeled arguments
  - Gains relative to local model on PropBank correct parses 88.4 \rightarrow 91.2 (24\% error reduction); gains on automatic parses 78.2 \rightarrow 80.0

- Also tree CRFs [Cohn & Brunson] have been used
Semantic roles: joint models boost results [Toutanova et al. 2005]

Accuracies of local and joint models on core arguments

Error reduction from best published result: 44.6% on Integrated 52% on Classification
System Properties

- Features
  - Most modern systems use the standard set of Gildea, Pradhan, and Surdeanu features listed above
  - *Lots of features important for building a good system*

- Learning Methods
  - SNoW, MaxEnt, AdaBoost, SVM, CRFs, etc.
  - *The choice of learning algorithms is less important.*
System Properties – Continued

- Syntactic Information
  - Charniak’s parser, Collins’ parser, clouser, chunker, etc.
  - Top systems use Charniak’s parser or some mixture
    - Quality of syntactic information is important

- System/Information Combination
  - Greedy, Re-ranking, Stacking, ILP inference
    - Combination of systems or syntactic information is a good strategy to reduce the influence of incorrect syntactic information!
Per Argument Performance
CoNLL-05 Results on WSJ-Test

- **Core Arguments** (Freq. ~70%)

<table>
<thead>
<tr>
<th></th>
<th>Best F$_1$</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0</td>
<td>88.31</td>
<td>25.58%</td>
</tr>
<tr>
<td>A1</td>
<td>79.91</td>
<td>35.36%</td>
</tr>
<tr>
<td>A2</td>
<td>70.26</td>
<td>8.26%</td>
</tr>
<tr>
<td>A3</td>
<td>65.26</td>
<td>1.39%</td>
</tr>
<tr>
<td>A4</td>
<td>77.25</td>
<td>1.09%</td>
</tr>
</tbody>
</table>

- **Adjuncts** (Freq. ~30%)

<table>
<thead>
<tr>
<th></th>
<th>Best F$_1$</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMP</td>
<td>78.21</td>
<td>6.86%</td>
</tr>
<tr>
<td>ADV</td>
<td>59.73</td>
<td>3.46%</td>
</tr>
<tr>
<td>DIS</td>
<td>80.45</td>
<td>2.05%</td>
</tr>
<tr>
<td>MNR</td>
<td>59.22</td>
<td>2.67%</td>
</tr>
<tr>
<td>LOC</td>
<td>60.99</td>
<td>2.48%</td>
</tr>
<tr>
<td>MOD</td>
<td>98.47</td>
<td>3.83%</td>
</tr>
<tr>
<td>CAU</td>
<td>64.62</td>
<td>0.50%</td>
</tr>
<tr>
<td>NEG</td>
<td>98.91</td>
<td>1.36%</td>
</tr>
</tbody>
</table>

Arguments that need to be improved

Data from Carreras&Màrquez’s slides (CoNLL 2005)
Summary

- Semantic role labeling
  - An important attempt at a general approach to shallow semantic extraction
- Relatively successful in terms of approximating
  - Human FrameNet labels
  - Human PropBank labels
- Are these good for anything?
  - We don’t know for sure yet