

---

# Semantic Role Labeling

## CS 224N

### Bill MacCartney

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 **Chris Manning**.

# 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

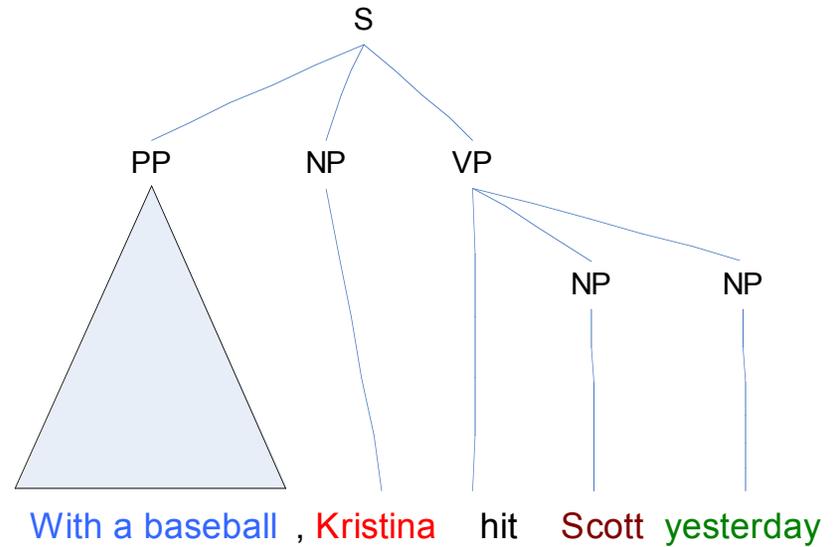
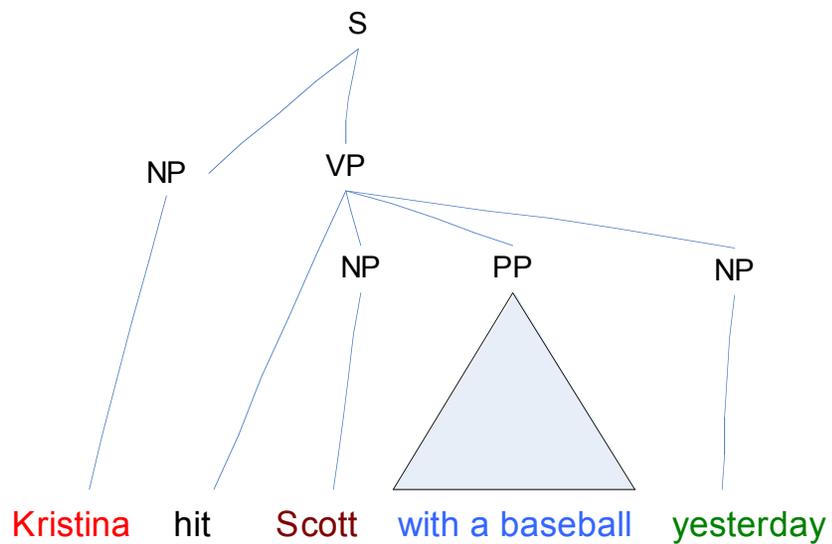
Patient, Thing hit

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

---

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

# Some typical semantic roles

---

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

# What is SRL good for?

## *Question Answering*

---

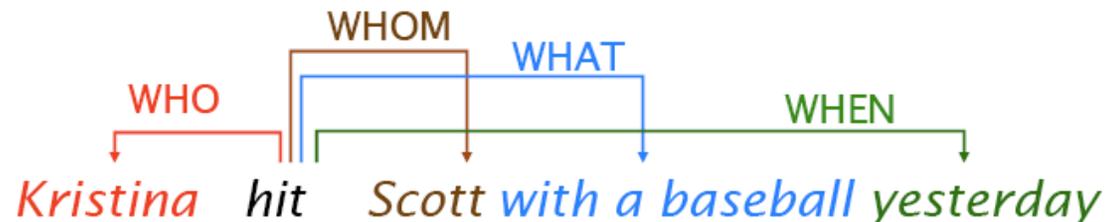
Q: *What was the name of the first computer system that defeated Kasparov?*

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:



- **Who** hit Scott with a baseball?
- **Whom** did Kristina hit with a baseball?
- **What** did Kristina hit Scott with?
- **When** did Kristina hit Scott with a baseball?

# 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

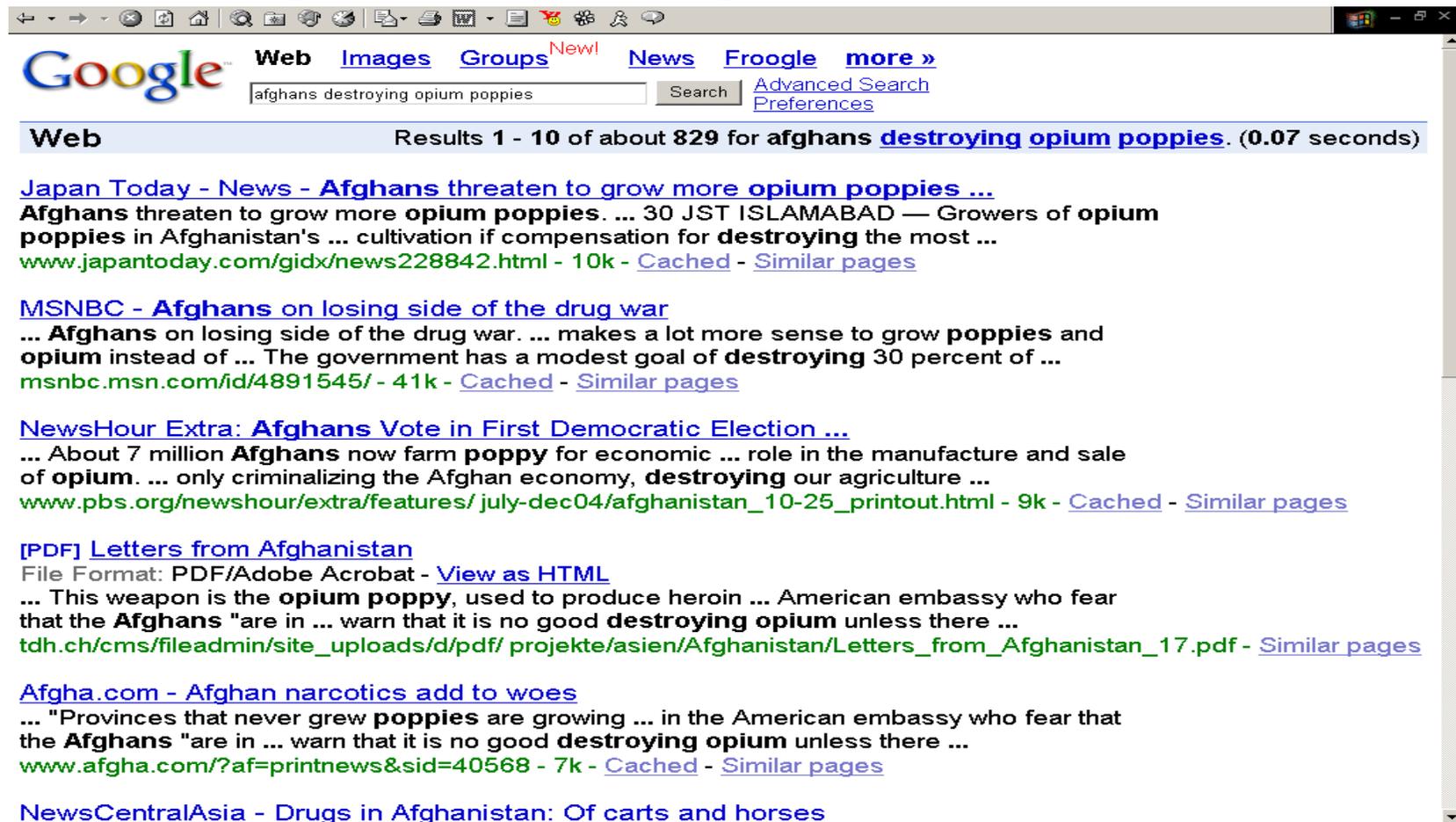
- Predicates and Heads of Roles summarize content

- Information Extraction

- SRL can be used to construct useful rules for IE

# Application: Semantically precise search

Query: *afghans destroying opium poppies*



The screenshot shows a Google search interface with the query "afghans destroying opium poppies" entered in the search box. The search results are displayed under the "Web" tab, showing 10 results out of approximately 829. The first result is from Japan Today, titled "Afghans threaten to grow more opium poppies ...". The second result is from MSNBC, titled "Afghans on losing side of the drug war". The third result is from NewsHour Extra, titled "Afghans Vote in First Democratic Election ...". The fourth result is a PDF document titled "Letters from Afghanistan". The fifth result is from Afgha.com, titled "Afghan narcotics add to woes". The sixth result is from NewsCentralAsia, titled "Drugs in Afghanistan: Of carts and horses".

Google Web Images Groups <sup>New!</sup> News Froogle more »

afghans destroying opium poppies Search [Advanced Search](#) [Preferences](#)

Web 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](http://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/4891545/](http://msnbc.msn.com/id/4891545/) - 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](http://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](http://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](#)  
... "Provinces 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&sid=40568](http://www.afgha.com/?af=printnews&sid=40568) - 7k - [Cached](#) - [Similar pages](#)

[NewsCentralAsia - Drugs in Afghanistan: Of carts and horses](#)



# 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: Agreeer

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

[<sub>A0</sub> *Kristina*] **hit** [<sub>A1</sub> *Scott*] [<sub>A2</sub> *with a baseball*] *yesterday*.

AM-TMP  
Time

- look.02 “seeming”

- ❖ A0: seemer; A1: seemed like; A2: seemed to

[<sub>A0</sub> *It*] **looked** [<sub>A2</sub> *to her*] *like* [<sub>A1</sub> *he deserved this*].

- deserve.01 “deserve”

- ❖ A0: deserving entity; A1: thing deserved;  
A2: in-exchange-for

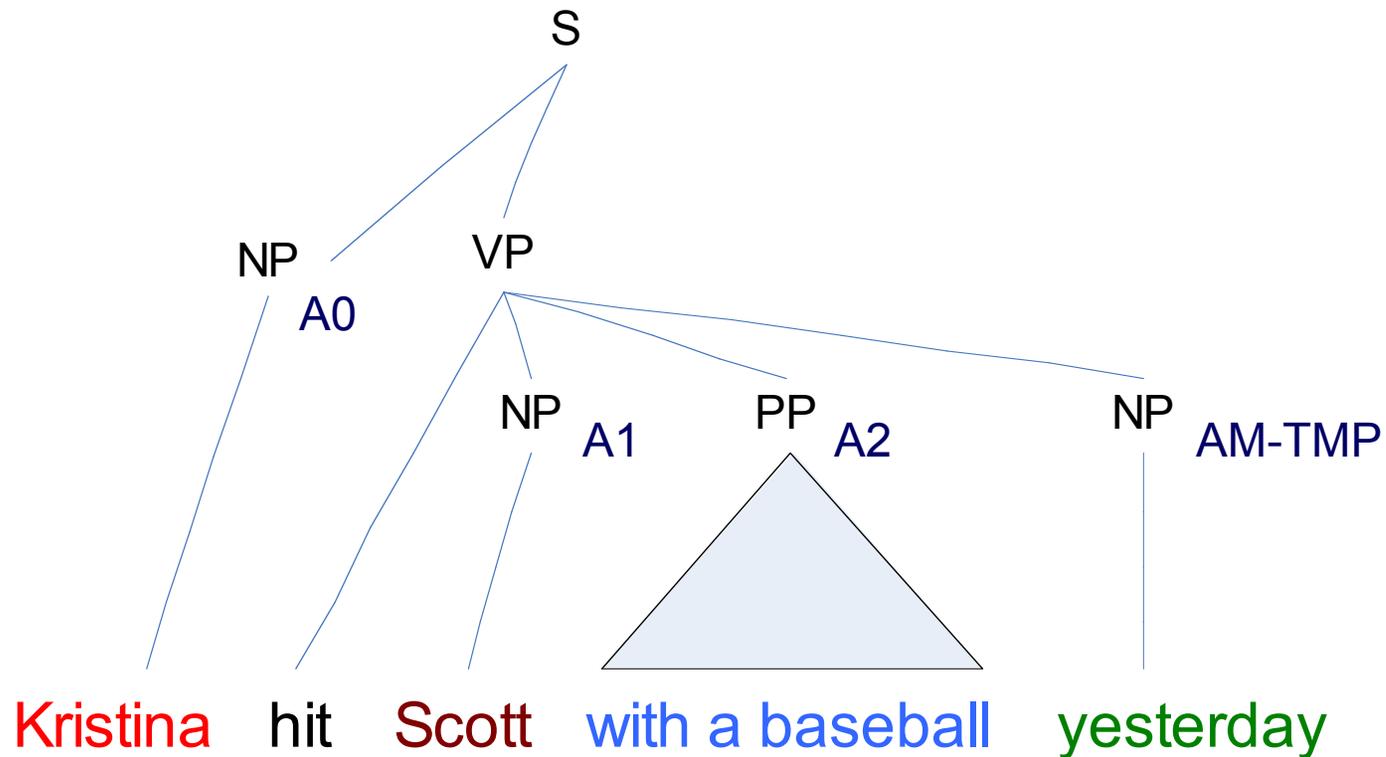
*It looked to her like* [<sub>A0</sub> *he*] **deserved** [<sub>A1</sub> *this*].

Proposition:  
A sentence and  
a target verb

# Proposition Bank (PropBank)

## Add a Semantic Layer

---

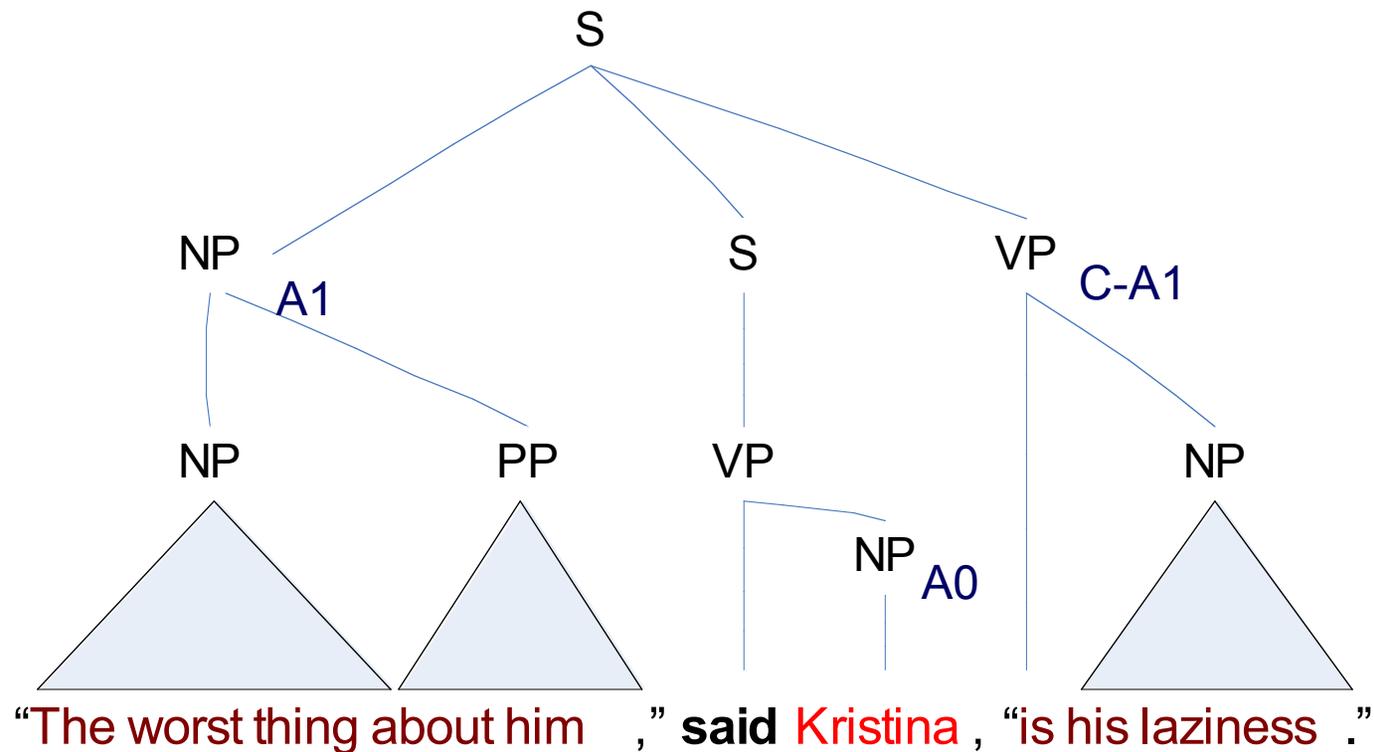


[<sub>A0</sub> *Kristina*] *hit* [<sub>A1</sub> *Scott*] [<sub>A2</sub> *with a basebal*] [<sub>AM-TMP</sub> *yesterday*].

# Proposition Bank (PropBank)

## Add a Semantic Layer – Continued

---



[<sub>A1</sub> *The worst thing about him*] said [<sub>A0</sub> *Kristina*] [<sub>C-A1</sub> *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

<http://www.lsi.upc.es/~srlconll/soft.html>

# 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")

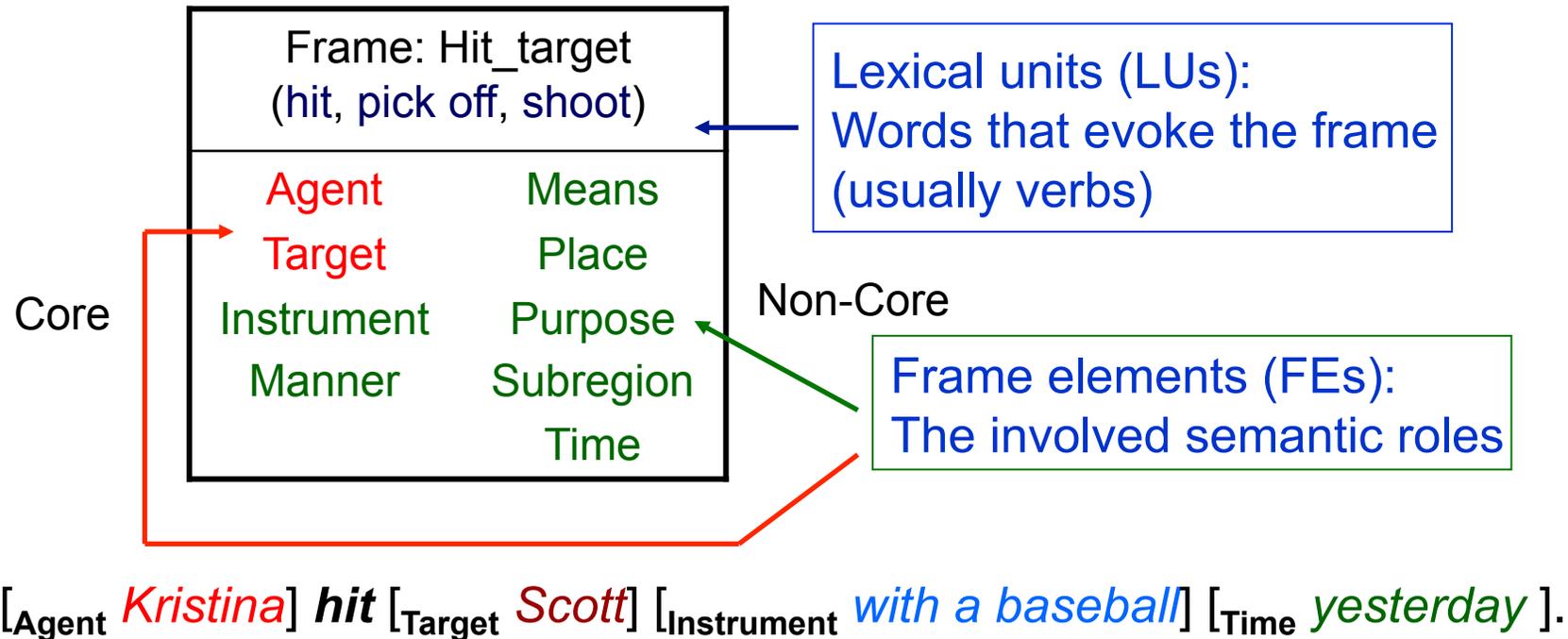
He	-	(AO*)	(AO*)	*	*
is	-	*	*	*	*
also	-	*	*	*	*
accused	-	*	*	*	*
of	-	*	*	*	*
lying	lie	(V*)	*	*	*
under	-	(AM-LOC*)	*	*	*
oath	-	*	*	*	*
and	-	*	*	*	*
of	-	*	*	*	*
leaking	leak	*	(V*)	*	*
information	-	*	(A1*)	(A1*)	*
obtained	obtain	*	*	(V*)	*
from	-	*	*	*	*
a	-	*	*	(A2*)	(A1*)
wiretap	-	*	*	*	*
he	-	*	*	*	(AO*)
supervised	supervise	*	*	*	(V*)
.	-	*	*	*	*

# 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**]

# FrameNet [Fillmore et al. 01]



<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

<b>VERBS:</b>	dwindle	move	soar	escalation	shift
advance	edge	mushroom	swell	explosion	tumble
climb	explode	plummet	swing	fall	
decline	fall	reach	triple	fluctuation	<b>ADVERBS:</b>
decrease	fluctuate	rise	tumble	gain	increasingly
diminish	gain	rocket		growth	
dip	grow	shift	<b>NOUNS:</b>	hike	
double	increase	skyrocket	decline	increase	
drop	jump	slide	decrease	rise	

# Roles in this frame

---

<b>Core Roles</b>	
ATTRIBUTE	The ATTRIBUTE is a scalar property that the ITEM possesses.
DIFFERENCE	The distance by which an ITEM changes its position on the scale.
FINAL_STATE	A description that presents the ITEM's state after the change in the ATTRIBUTE's value as an independent predication.
FINAL_VALUE	The position on the scale where the ITEM ends up.
INITIAL_STATE	A description that presents the ITEM's state before the change in the ATTRIBUTE's value as an independent predication.
INITIAL_VALUE	The initial position on the scale from which the ITEM moves away.
ITEM	The entity that has a position on the scale.
VALUE_RANGE	A portion of the scale, typically identified by its end points, along which the values of the ATTRIBUTE fluctuate.
<b>Some Non-Core Roles</b>	
DURATION	The length of time over which the change takes place.
SPEED	The rate of change of the VALUE.
GROUP	The GROUP in which an ITEM changes the value of an ATTRIBUTE in a specified way.

# Examples / Quiz question

---

## Examples:

- [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].

## Quiz: give the roles for the items with blanks:

- a steady *increase* [            from 9.5] [            to 14.3] [            in dividends]
- a [            5%] [            dividend] *increase*

- (A) INITIAL\_STATE, FINAL\_STATE, ATTRIBUTE; DIFFERENCE, ATTRIBUTE
- (B) INITIAL\_VALUE, FINAL\_VALUE, ATTRIBUTE; VALUE\_RANGE, ATTRIBUTE
- (C) INITIAL\_VALUE, FINAL\_VALUE, ITEM; DIFFERENCE, ITEM
- (D) none of the above

# 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 1

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].

PROPBANK 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:

[Goods A car] was *bought* [Buyer by Chuck].

[Goods A car] was *sold* [Buyer to Chuck] [Seller by Jerry].

[Buyer Chuck] was *sold* [Goods a car] [Seller by Jerry].

PROPBANK ANNOTATION:

[Arg1 A car] was *bought* [Arg0 by Chuck].

[Arg1 A car] was *sold* [Arg2 to Chuck] [Arg0 by Jerry].

[Arg2 Chuck] was *sold* [Arg1 a car] [Arg0 by Jerry].

# Information Extraction versus Semantic Role Labeling

---

<b>Characteristic</b>	<b>IE</b>	<b>SRL</b>
Coverage	narrow	broad
Depth of semantics	shallow	shallow
Directly connected to application	sometimes	no

- The goal of SRL is to provide a general capability for semantic relation and argument identification, not custom ones for particular applications

# 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:**  $2^{\{1,2,\dots,m\}} \mapsto \{NONE, ARG\}$ 
  - 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:**  $2^{\{1,2,\dots,m\}} \mapsto L \setminus \{NONE\}$ 
  - 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: [<sub>A0</sub> The queen] **broke** [<sub>A1</sub> the window] [<sub>AM-TMP</sub> yesterday]

Guess: [<sub>A0</sub> The queen] broke the [<sub>A1</sub> window] [<sub>AM-LOC</sub> yesterday]

Correct	Guess
{The queen} → A0	{The queen} → A0
{the window} → A1	{window} → A1
{yesterday} → AM-TMP	{yesterday} → AM-LOC
all other → NONE	all other → NONE

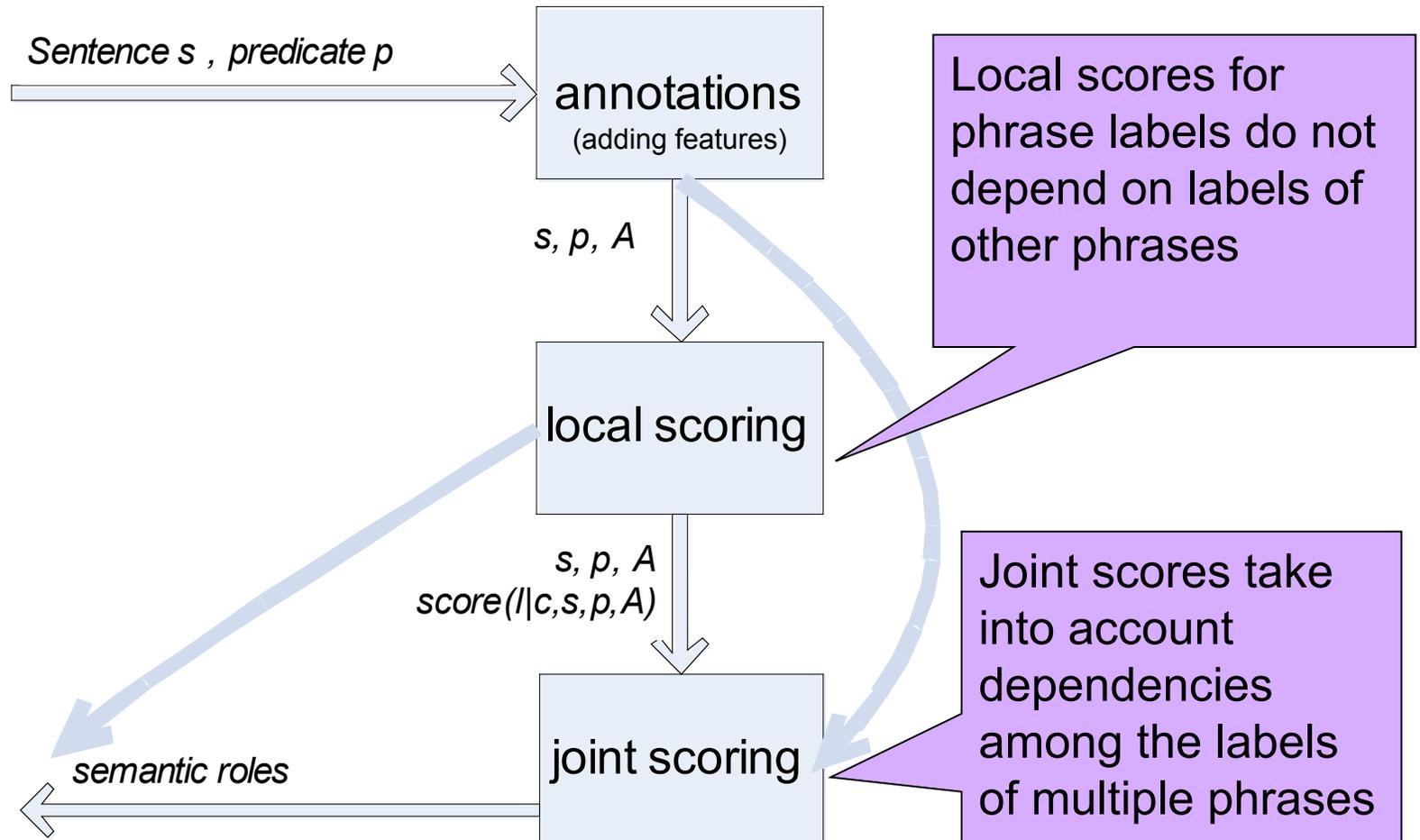
- Precision, Recall, F-Measure { $tp=1, fp=2, fn=2$ }  $p=r=f=1/3$
- Measures for subtasks
  - Identification (Precision, Recall, F-measure) { $tp=2, fp=1, fn=1$ }  $p=r=f=2/3$
  - Classification (Accuracy)  $acc = .5$  (labeling of correctly identified phrases)
  - Core arguments (Precision, Recall, 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

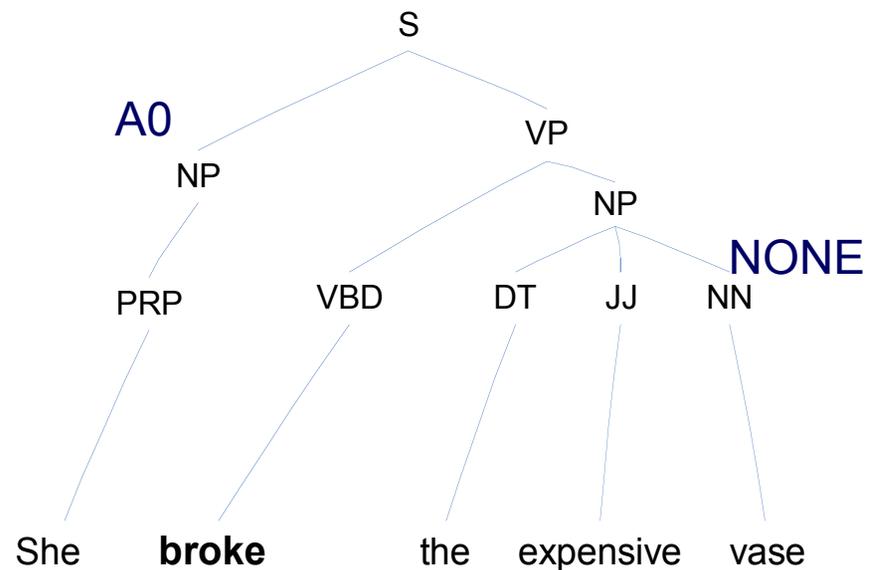


# SRL architecture:

## 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..*

[<sub>NP</sub>Yesterday] , [<sub>NP</sub>Kristina] [<sub>VP</sub>hit] [<sub>NP</sub>Scott] [<sub>PP</sub>with] [<sub>NP</sub>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

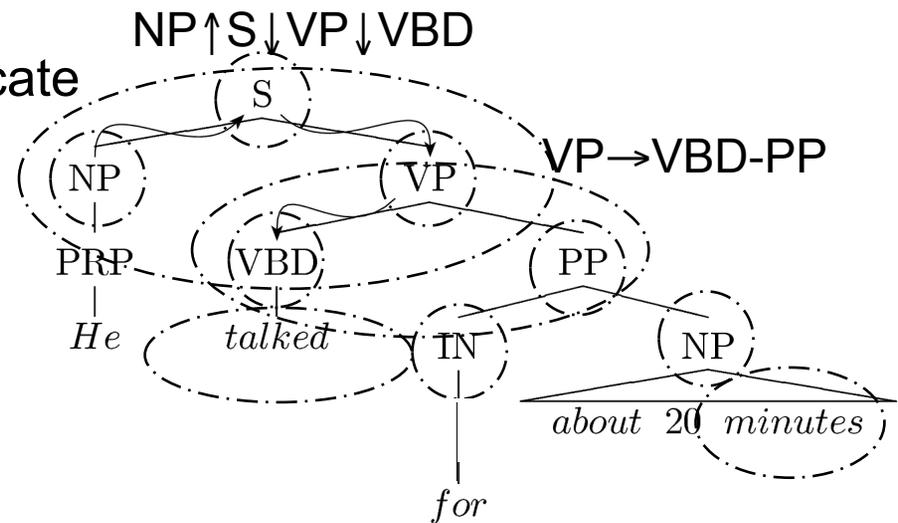
# SRL Parsing Algorithm

---

- 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



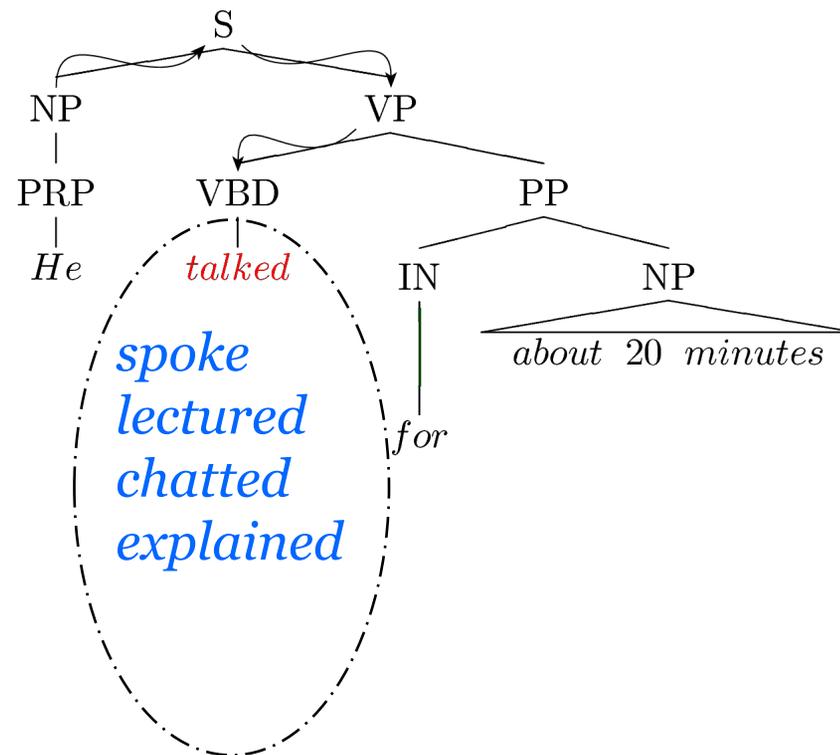
# Pradhan et al. (2004) Features

---

- 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

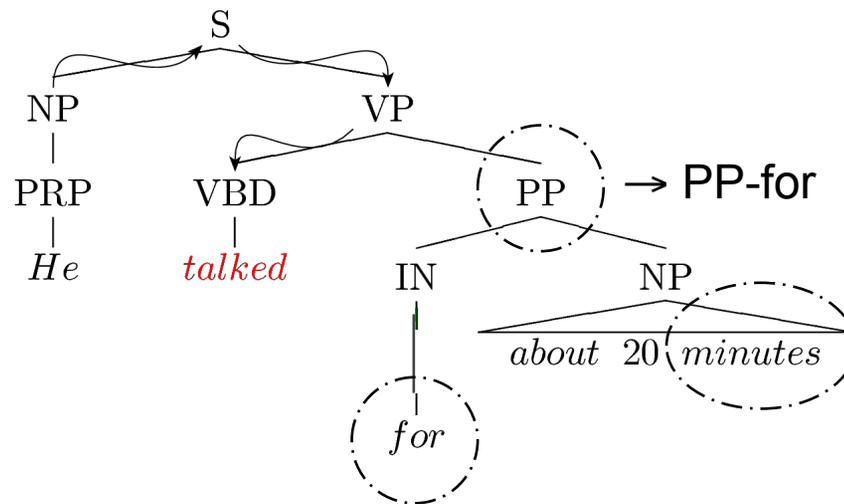
# Predicate cluster, automatic or WordNet

---



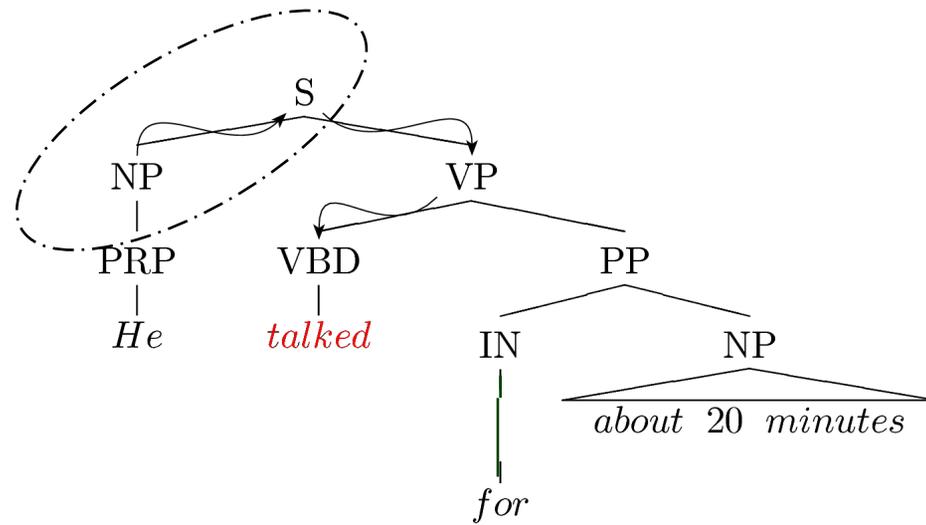
# Noun Head and POS of PP

---



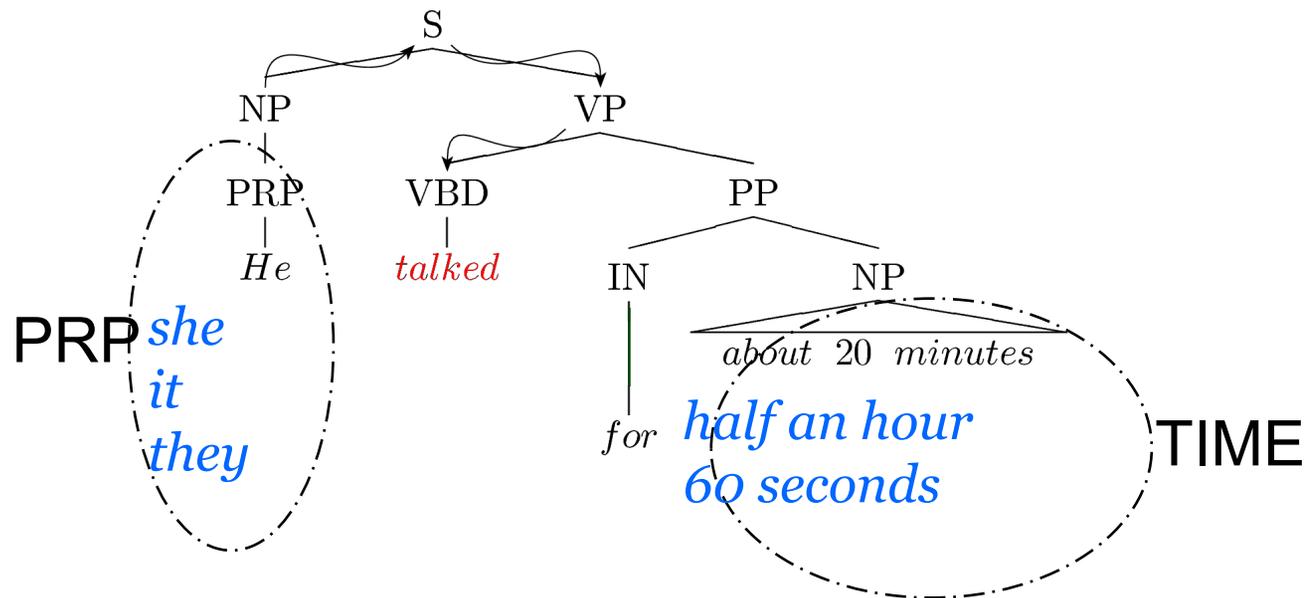
# Partial Path

---



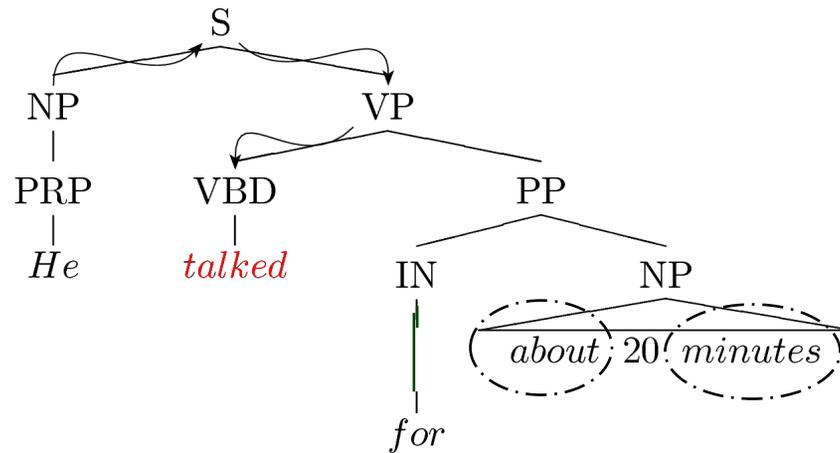
# Named Entities and Head Word POS

[Surdeanu et al., 2003]



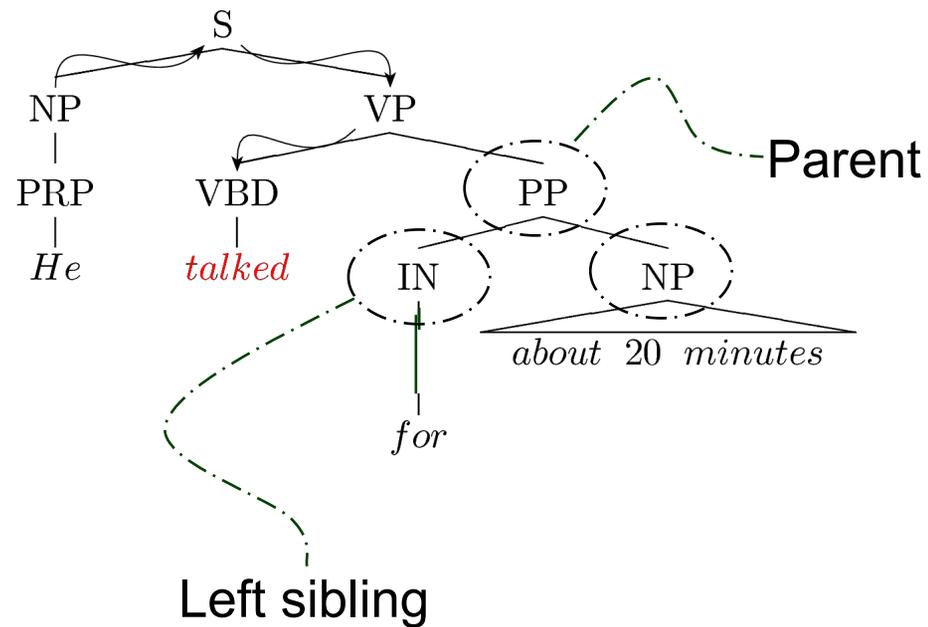
# First and Last Word and POS

---



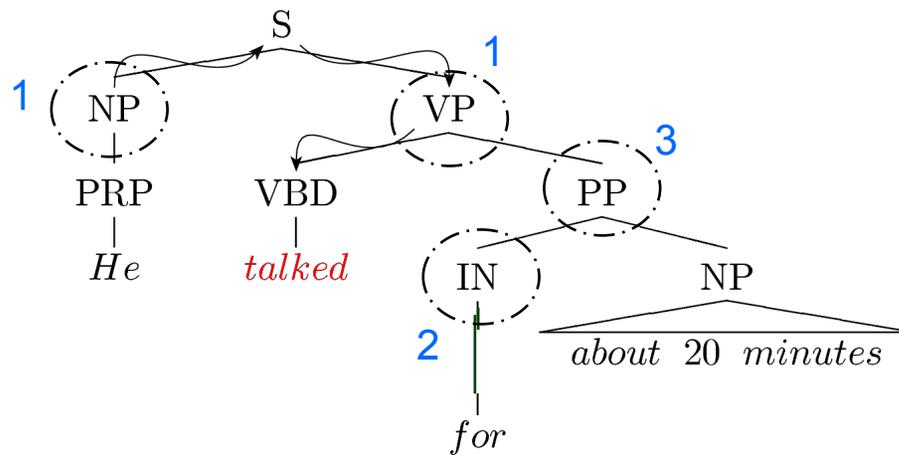
# Parent and Sibling features

---



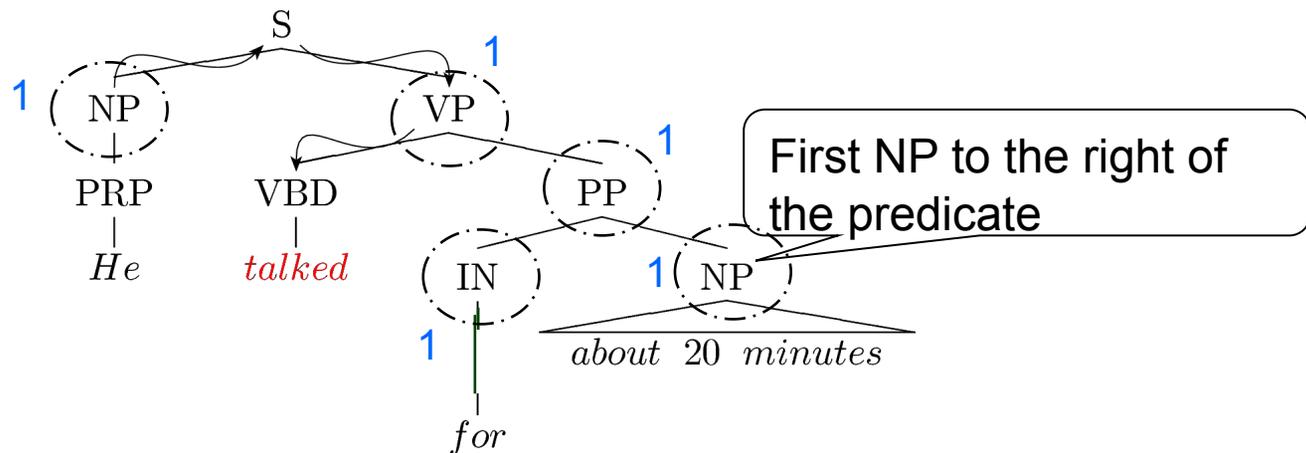
# Constituent tree distance

---



# Ordinal constituent position

---



## Temporal Cue Words (~50)

---

time      years;ago

recently    night

days      hour

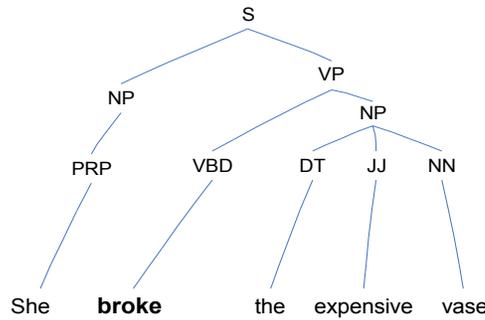
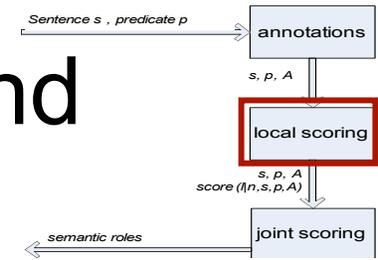
end        decade

period     late

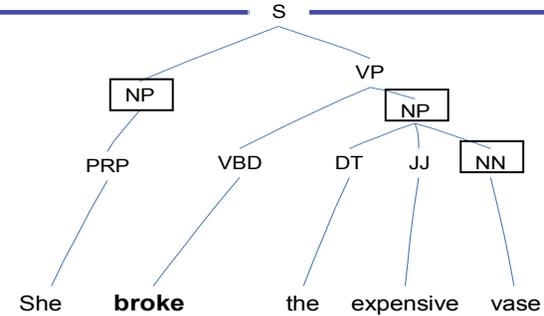
# Previous 2 classifications

---

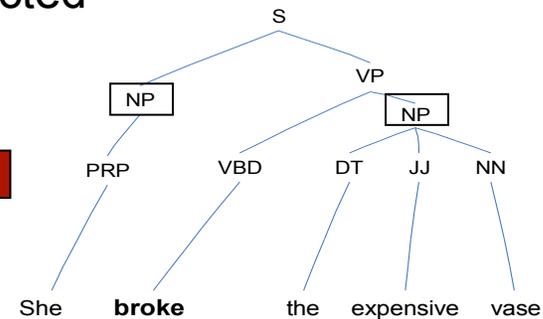
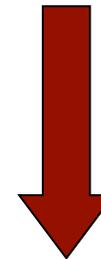
# Combining Identification and Classification Models



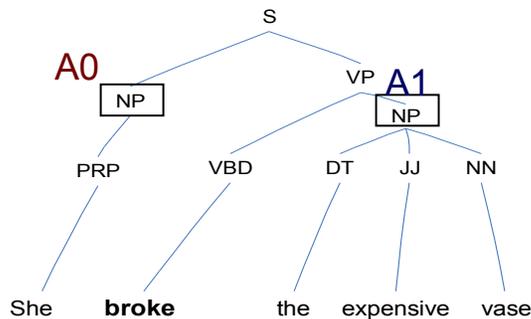
**Step 1. Pruning.**  
Using a hand-specified filter.



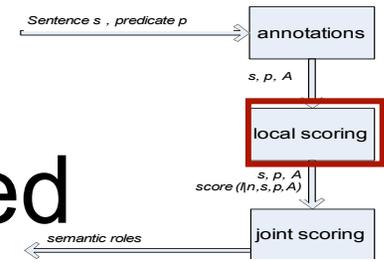
**Step 2. Identification.**  
Identification model (filters out candidates with high probability of NONE)



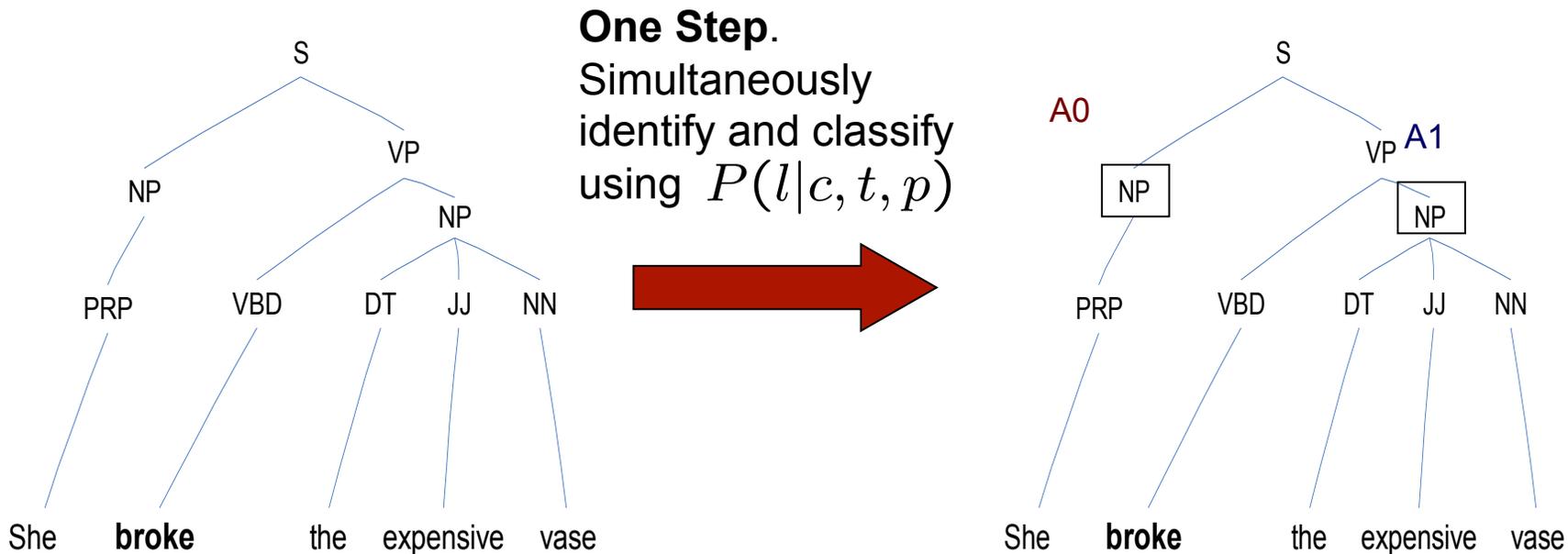
**Step 3. Classification.**  
Classification model assigns one of the argument labels to selected nodes (or sometimes possibly NONE)



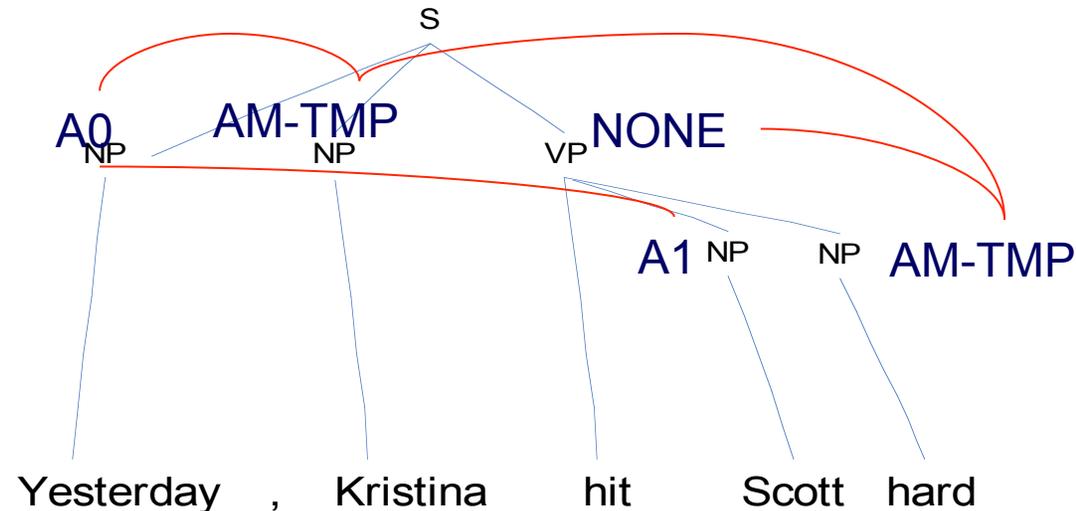
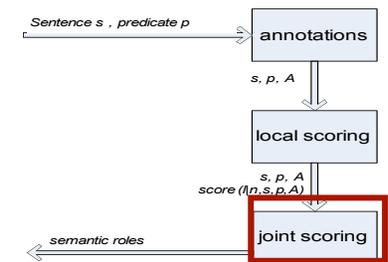
# Combining Identification and Classification Models – Continued



$$-P(l|c, t, p) = P_{ID}(Id(l)|\Phi(c, t, p)) * P_{CLS}(l|Id(l), \Phi(c, t, p))$$
 or
 
$$-P(l|c, t, p) = P(l|\Phi(c, t, p))$$



# Joint Scoring Models

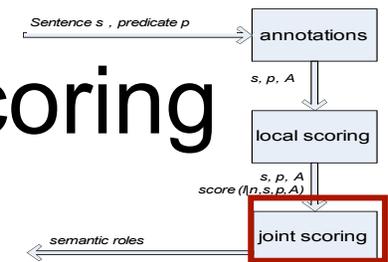


- These models have scores for a whole labeling of a tree (not just individual labels)

- Encode some dependencies among the labels of different nodes

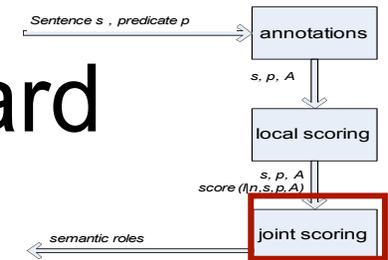
$$P_{JOINT}(l_1, \dots, l_n | n, t, p) \neq \prod_i P(l_i | n_i, t, p)$$

# 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

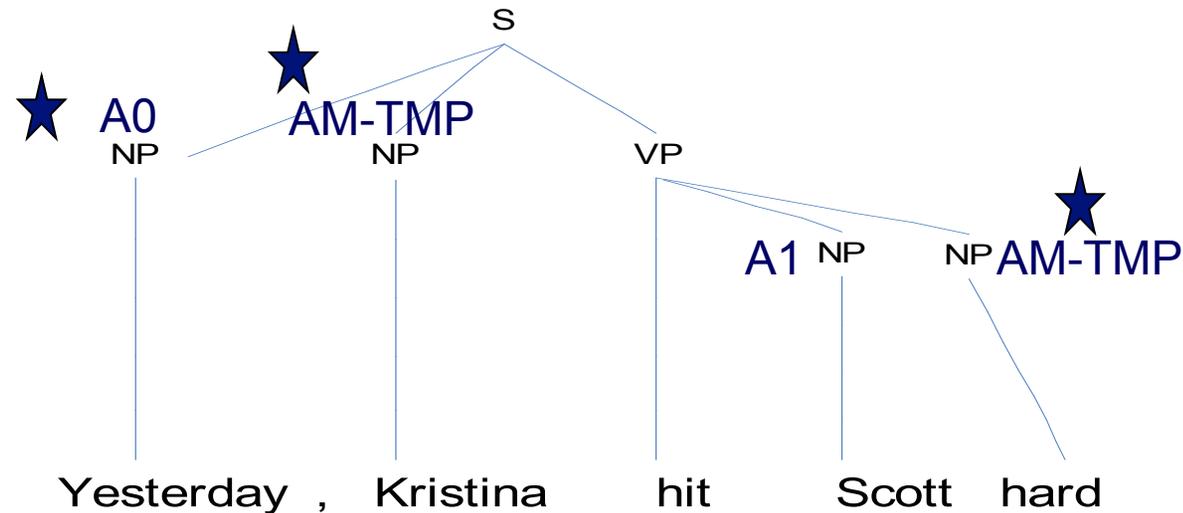
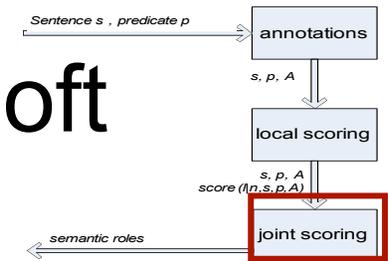


- **Constraint 1: Argument phrases do not overlap**

*By [A<sub>1</sub> working [A<sub>1</sub> hard ], he] **said** , you can achieve a lot.*

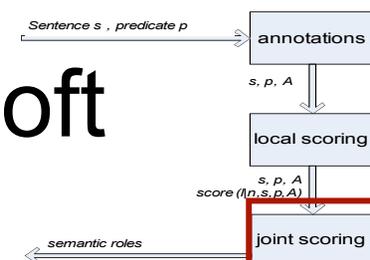
- Pradhan et al. (04) – greedy search for a best set of non-overlapping arguments
  - Toutanova et al. (05) – exact search for the best set of non-overlapping arguments (dynamic programming, linear in the size of the tree)
  - Punyakanok et al. (05) – exact search for best non-overlapping arguments using 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:

$$P(\{A0, AM_{TMP}, A1, AM_{TMP}\} | hit)$$

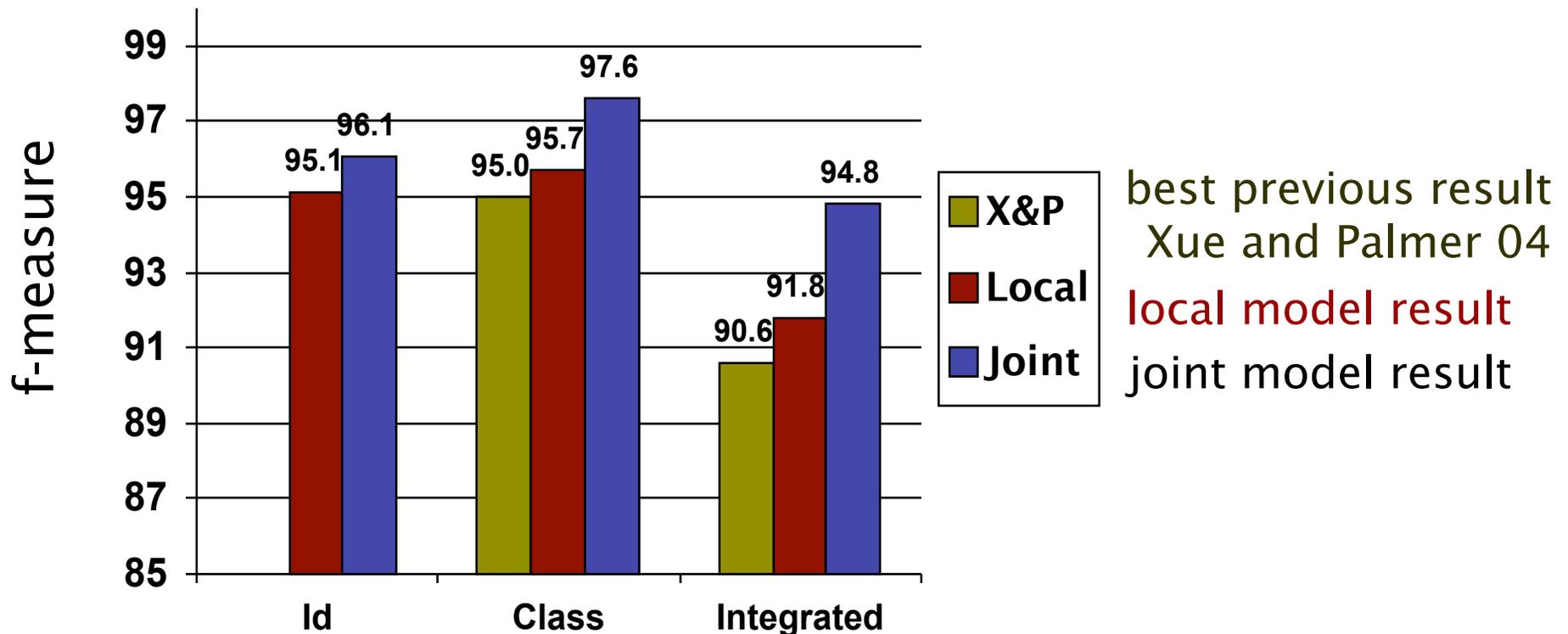
- Gains relative to local model 59.2 → 62.9 FrameNet automatic parses
- Pradhan et al. (04) – a language model on argument label sequences (with the predicate included)

$$P(A0, AM_{TMP}, hit, A1, AM_{TMP})$$

- Small gains relative to local model for a baseline system 88.0 → 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 → 91.2 (24% error reduction); gains on automatic parses 78.2 → 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, clauser, 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%)

	Best F <sub>1</sub>	Freq.
A0	88.31	25.58%
A1	79.91	35.36%
A2	70.26	8.26%
A3	65.26	1.39%
A4	77.25	1.09%

- Adjuncts (Freq. ~30%)

	Best F <sub>1</sub>	Freq.
TMP	78.21	6.86%
ADV	59.73	3.46%
DIS	80.45	2.05%
MNR	59.22	2.67%
LOC	60.99	2.48%
MOD	98.47	3.83%
CAU	64.62	0.50%
NEG	98.91	1.36%

Arguments that need to be improved

# 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