CS114 Lecture 18
Semantic Roles

March 31, 2014

Professor Meteer

Thanks for Jurafsky & Martin & Prof. Pustejovksy for slides

What are semantic roles and what is their history?

- A lot of forms of traditional grammar (Sanskrit, Japanese, …) analyze in terms of a rich array of semantically potent case ending or particles
  - They're kind of like semantic roles
- The idea resurfaces in modern generative grammar in work of Charles ("Chuck") Fillmore, who calls them Case Roles (Fillmore, 1968, The Case for Case).
  - They're quickly renamed to other words, but various:
    - Semantic roles
    - Thematic roles
    - Theta roles
    - A predicate and its semantic roles are often taken together as an **argument structure**

Okay, but what *are* they?

- An event is expressed by a predicate and various other dependents
- The claim of a theory of semantic roles is that these other dependents can be usefully classified into a small set of semantically contentful classes
  - And that these classes are useful for explaining lots of things

Common semantic roles

- Agent: initiator or doer in the event
  - **Sue** killed **the rat**.
- Patient: affected entity in the event; undergoes the action
- Theme: object in the event undergoing a change of state or location, or of which location is predicated
  - **The ice** melted
- Experiencer: feels or perceive the event
  - **Bill** likes **pizza**.
- Stimulus: the thing that is felt or perceived

Slide from Chris Manning

Common semantic roles

- **Goal**:
  - Bill ran <u>to Copley Square</u>.

- **Recipient** (may or may not be distinguished from Goal):
  - Bill gave the book <u>to Mary</u>.

- **Benefactive** (may be grouped with Recipient):
  - Bill cooked dinner <u>for Mary</u>.

- **Source**:
  - Bill took a pencil <u>from the pile</u>.

- **Instrument**:
  - Bill ate the burrito <u>with a plastic spork</u>.

- **Location**:
  - Bill sits <u>under the tree</u> on Wednesdays

Common semantic roles

Try for yourself!
1. The submarine sank a troop ship.
2. Doris hid the money in the flowerpot.
3. Emma noticed the stain.
4. We crossed the street.
5. The boys climbed the wall.
6. The chef cooked a great meal.
7. The computer pinpointed the error.
8. A mad bull damaged the fence on Jack's farm.
9. The company wrote me a letter.
10. Jack opened the lock with a paper clip.

Linking of thematic roles to syntactic positions

- John opened the door
- AGENT   THEME
- The door was opened by John
- THEME   AGENT
- The door opened
- THEME
- John opened the door with the key
- AGENT   THEME   INSTRUMENT

Deeper Semantics

- From the WSJ…
  - He melted her reserve with a husky-voiced paean to her eyes.
  - If we label the constituents He and her reserve as the Melter and Melted, then those labels lose any meaning they might have had.
  - If we make them Agent and Theme then we can do more inference.

Problems

- What exactly is a role?

- What's the right set of roles?

- Are such roles universals?

- Are these roles atomic?
  - I.e. Agents
    - Animate, Volitional, Direct causers, etc

- Can we automatically label syntactic constituents with thematic roles?

Syntactic Variations

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

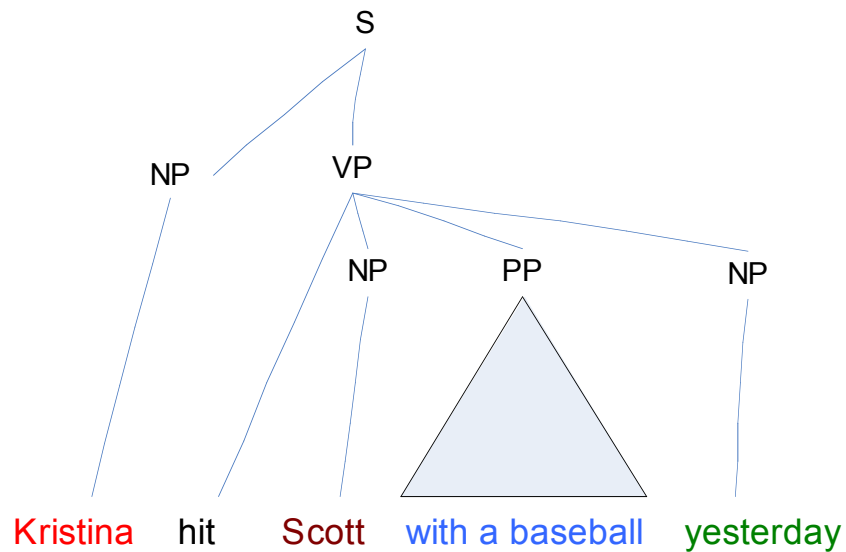*Kristina hit Scott with a baseball yesterday*

| Agent, hitter | Thing hit | Instrument | Temporal adjunct |

Syntactic Variations (as trees)
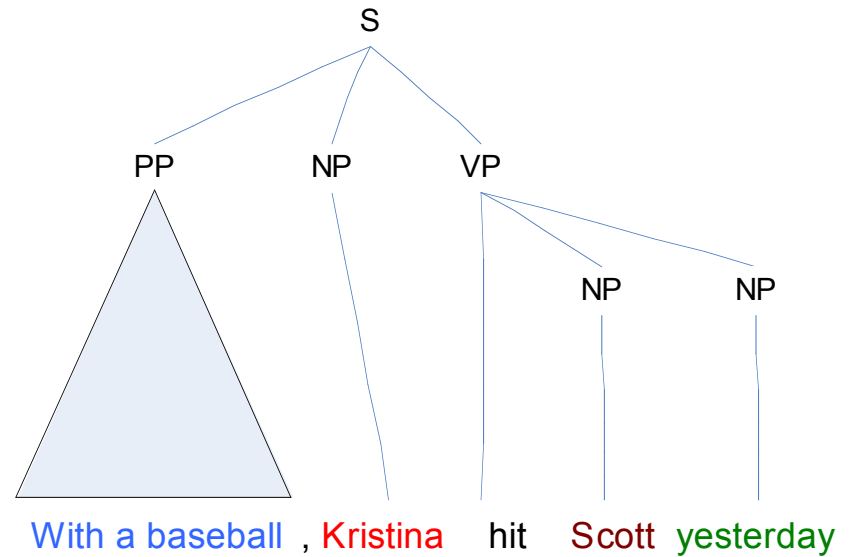
Semantic Role Labeling – *Giving Semantic Labels to Phrases*

- [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]

Why is SRL Important – *Applications*

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

- Machine Translation

  English (SVO)   Farsi (SOV)

  [AGENT The little boy]   [AGENT pesar koocholo] boy-little

  [PRED **kicked**]   [THEME toop germezi]   ball-red
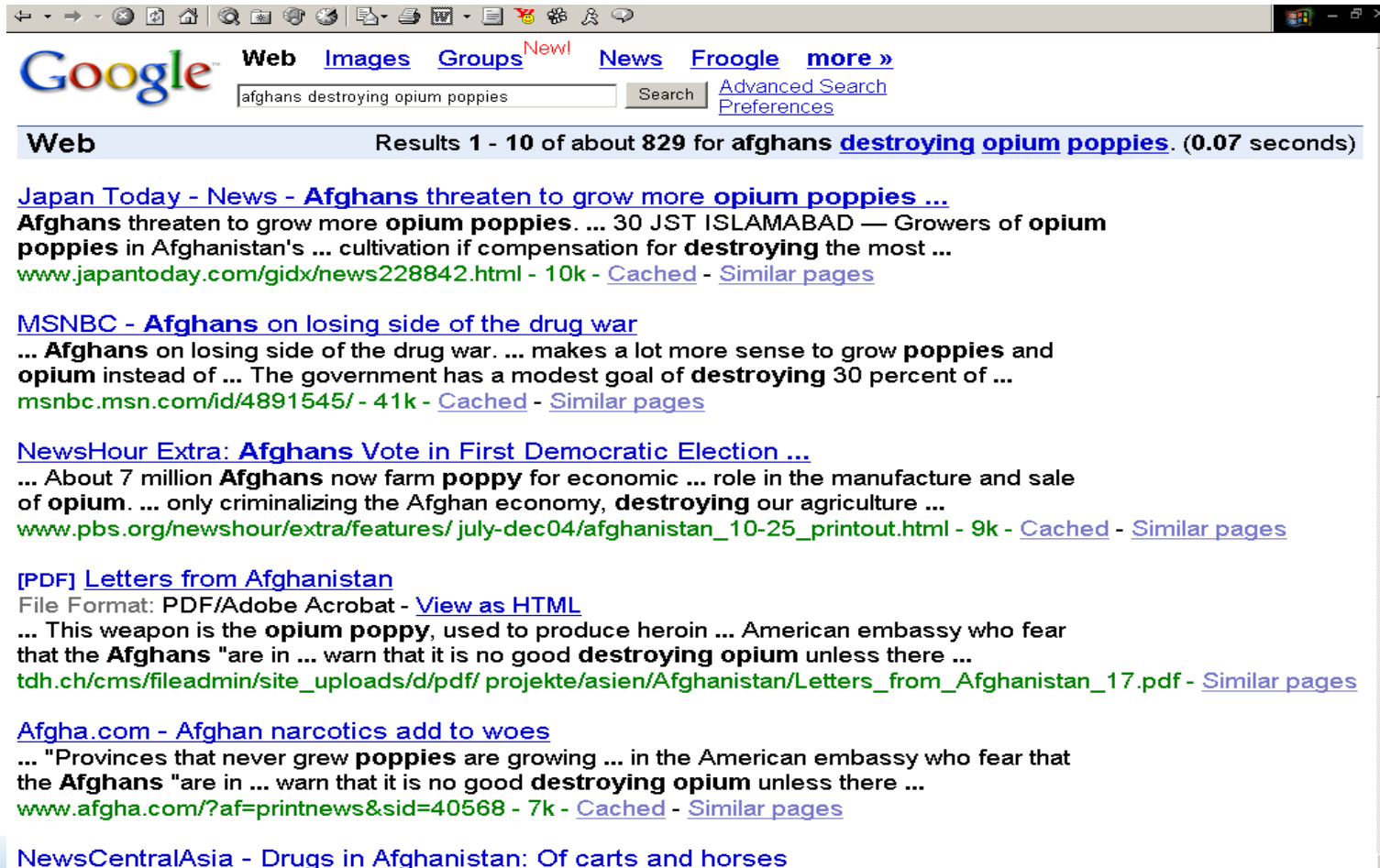
  [THEME the red ball]   [ARGM-MNR moqtam]   hard-adverb

  [ARGM-MNR hard]   [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*

Some History

- Minsky 74, Fillmore 1976: *frames* describe events or situations
  - Multiple participants, "props", and "conceptual roles"
- Levin 1993: verb class defined by sets of frames (meaning-preserving alternations) a verb appears in
  - {*break,shatter,..*}: *Glass X's easily; John Xed the glass, …*
  - *Cut* is different: *The window broke; *The window cut.*
- FrameNet, late '90s: based on Levin's work: large corpus of sentences annotated with *frames*
- PropBank: addresses tragic flaw in FrameNet corpus

Levin's Verb Classes

- Beth Levin analyzed thousands of verbs and defined hundreds of classes

- Underlying hypothesis:
  - verbal meaning determines syntactic realizations

- Examples
  Touch: kiss, sting
  Hit: Bash, hammer, tap
  Cut: chip, hack, scratch
  Break: back, split, tear.

- Conative
  Jean moved the table
  *Jean moved at the table

- Body-part possessor ascension
  Janet broke Bill's finger
  *Janet broke Bill on the finger

- Middle
  Bread cuts easily
  Cats touch easily

Alternation	Touch	Hit	Cut	Break
Conative	N	Y	Y	N
Body part possessor ascension	Y	Y	Y	N
Middle	N	N	Y	Y

Frames in Framenet

Figure 1:A subframe can inherit elements and semantic from it's parent

frame(TRANSPORTATION) frame_elements(MOVER(s), MEANS, PATH) scene(MOVER(s) move along PATH by MEANS)
frame(DRIVING) inherit (TRANSPORTATION) frame_elements(DRIVER (=MOVER), VEHICLE(=MEANS), RIDER(S) (= MOVER(S)), CARGO (=MOVER(S))) scenes(DRiVER starts VEHICLE, DRIVER controls VEHICLE. DRIVER stops VEHICLE)
frame(RIDING_I) inherit (TRANSPORTATION) frame_elements(RiDER(s) (=MOVER(S)), VEHICLE (=MEANS)) scenes(RiDER enters VEHICLE,VEHICLE carries RIDER along PATH,RIDER leaves VEHICLE)

Frame Element Groups

Figure 2: Examples of Frame Element Groups and Annotated Sentences

FEG	Annotated Example
D	*[D* Kate] **drove** *[P* home] in a stupor.
D, V	A pregnant woman lost her baby after she fainted as she waited for a bus and fell into the path of [V a lorry] **driven** *[D* by her uncle].
D,P	And that was why *[D* I] **drove** [P eastwards along Lake Geneva].
D, R, P	Now *[D* Van Checle] was **driving** *[R* his guest] *[P* back to the station].
D, V, P	[D CummingJ had a fascination with most forms of transport, **driving** *[V* his Rolls] at high speed *[P* around the streets of London].
D+R, P	*[D* WeJ **drive** *[P* home along miles of empty freeway].
V, P	Over the next 4 days, *[V* the Rolls Royces] will **drive** *[P* down to Plymouth], following the route of the railway.

FrameNet [Fillmore et al. 01]

Frame: Hit_target
(hit, pick off, shoot)

Core		Non-Core
Agent	Means	
Target	Place	
Instrument	Purpose	
Manner	Subregion	
	Time	

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

Frame elements (FEs):
The involved semantic roles
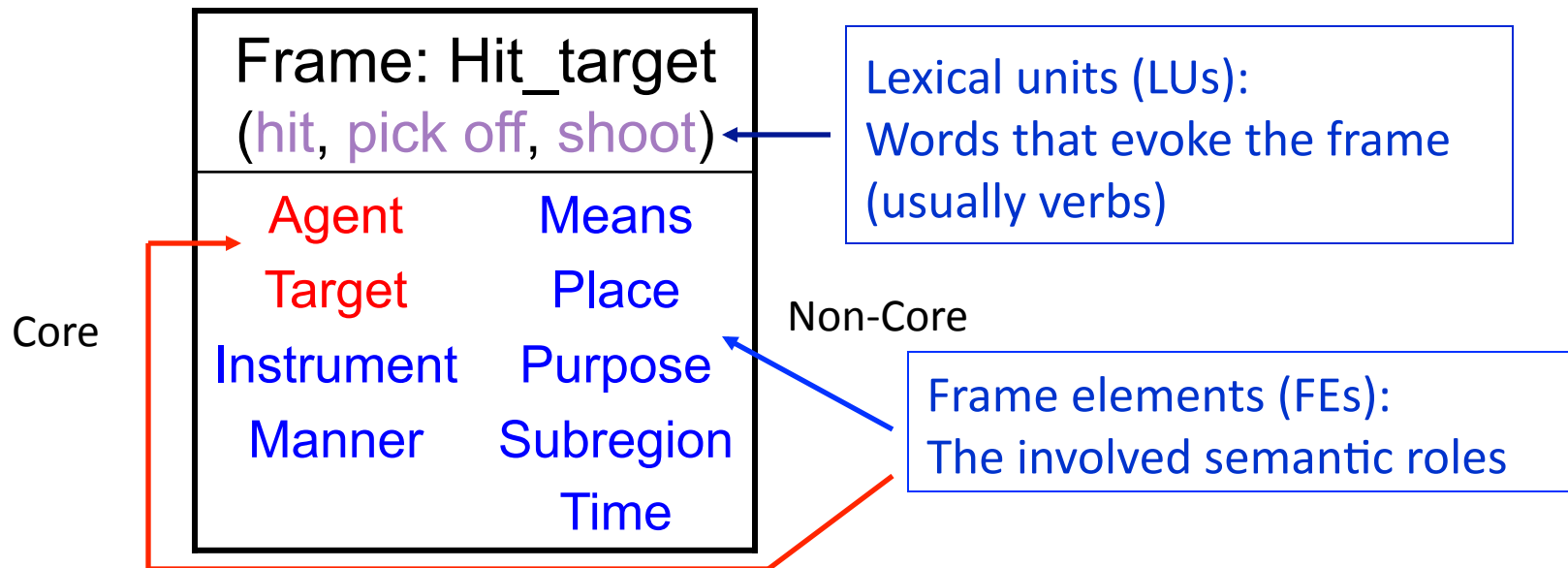
[**Agent** *Kristina*] **hit** [**Target** *Scott*] [**Instrument** *with a baseball*] [**Time** *yesterday*].

Methodology for FrameNet

1. Define a frame (eg DRIVING)
2. Find some sentences for that frame
3. Annotate them
4. If (remaining funding == 0) then exit;   else goto step 1.

- Corpora
  - FrameNet I – British National Corpus only
  - FrameNet II – LDC North American Newswire corpora
- Size
  - >8,900 lexical units, >625 frames, >135,000 sentences

http://framenet.icsi.berkeley.edu

Annotations in PropBank

- Based on Penn TreeBank
- Goal is to annotate *every tree* systematically
  - so statistics in the corpus are meaningful
- Like FrameNet, based on Levin's verb classes (via VerbNet)
- Generally more data-driven & bottom up
  - No level of abstraction beyond verb senses
  - Annotate every verb you see, whether or not it seems to be part of a frame

Some verb senses and "framesets" for propbank

Frameset: decline.01 "go down incrementally"

    Arg1: entity going down

    Arg2: amount gone down by, EXT

    Arg3: start point

    Arg4: end point

Ex: ... [$_{arg1}$ its net income] declining [$_{arg2\ EXT}$ 42%][$_{arg4}$ to $121 million dollars] [$_{argM\text{-}TMP}$ in the first 9 months of 1989].

Frameset: decline.02 "demure.reject"

    Arg0: agent

    Arg1: rejected thing

Ex: [$_{arg0}$ A spokesman] declined [$_{arg1}$ "trace" to elaborate]

FrameNet vs PropBank

- Framenet annotation

  [buyer Chuck] bought [goods a car] [seller from Jerry] [payment for $1000].

  [seller Jerry] sold [goods a car] to [buyer Chuck] [payment for $1000].

  [goods A car] was sold [buyer to Chuck] [seller by Jerry].

  [buyer Chuck] was sold [goods a car] [seller by Jerry].

- Propbank Annotation

  [Arg0 Chuck] bought [Arg1 a car] [Arg2 from Jerry] [Arg3 for $1000].

  [Arg0 Jerry] sold [Arg1 a car] to [Arg2 Chuck] [Arg3 for $1000].

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

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

Proposition Bank (PropBank) [Palmer et al. 05]

- Transfer sentences to propositions
  - Kristina hit Scott → hit(Kristina,Scott)

- Penn TreeBank → PropBank
  - Add a semantic layer on Penn TreeBank
  - Define a set of semantic roles for each verb
  - Each verb's roles are numbered

…[A0 the company] to … *offer* [A1 a 15% to 20% stake] [A2 to the public]
…[A0 Sotheby's] … *offered* [A2 the Dorrance heirs] [A1 a money-back guarantee]
…[A1 an amendment] *offered* [A0 by Rep. Peter DeFazio] …
…[A2 Subcontractors] will be *offered* [A1 a settlement] …

Proposition Bank (PropBank)
Define the Set of Semantic Roles

- It's difficult to define a general set of semantic roles for all types of predicates (verbs).

- PropBank defines semantic roles for each verb and sense in the frame files.

- The (core) arguments are labeled by numbers.
  - A0 – Agent; A1 – Patient or Theme
  - Other arguments – no consistent generalizations

- Adjunct-like arguments – *universal* to all verbs
  - AM-LOC, TMP, EXT, CAU, DIR, PNC, ADV, MNR, NEG, MOD, DIS

Proposition Bank (PropBank) Frame Files

- ## hit.01 "strike"
  - ❖ A0: agent, hitter; A1: thing hit;
    A2: instrument, thing hit by or with

  [A0 *Kristina*] **hit** [A1 *Scott*] [A2 *with a baseball*] *yesterday*.

  AM-TMP
  Time

- ## look.02 "seeming"
  - ❖ A0: seemer; A1: seemed like; A2: seemed to

  [A0 *It*] **looked** [A2 *to her*] like [A1 *he deserved this*].
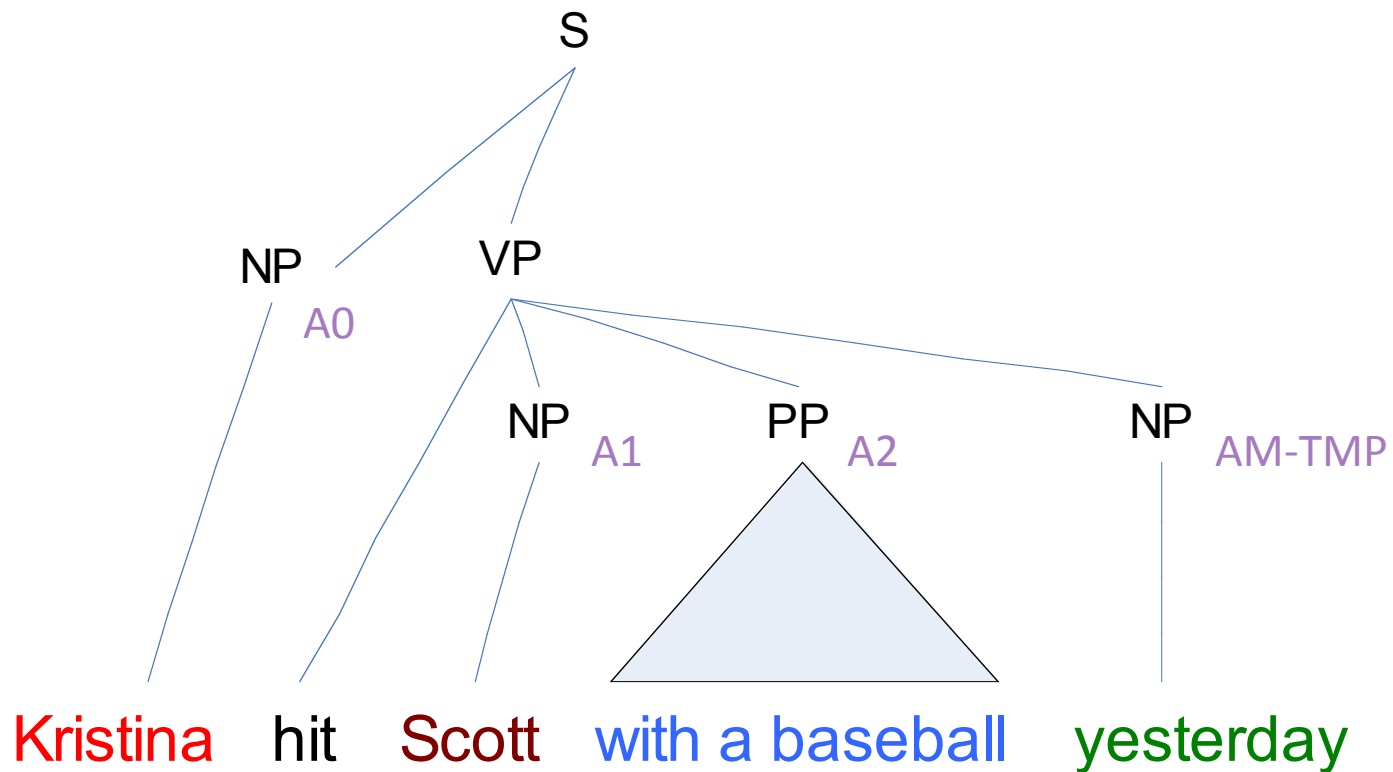
- ## deserve.01 "deserve"
  - ❖ A0: deserving entity; A1: thing deserved;
    A2: in-exchange-for

  It looked to her like [A0 *he*] **deserved** [A1 *this*].

  Proposition:
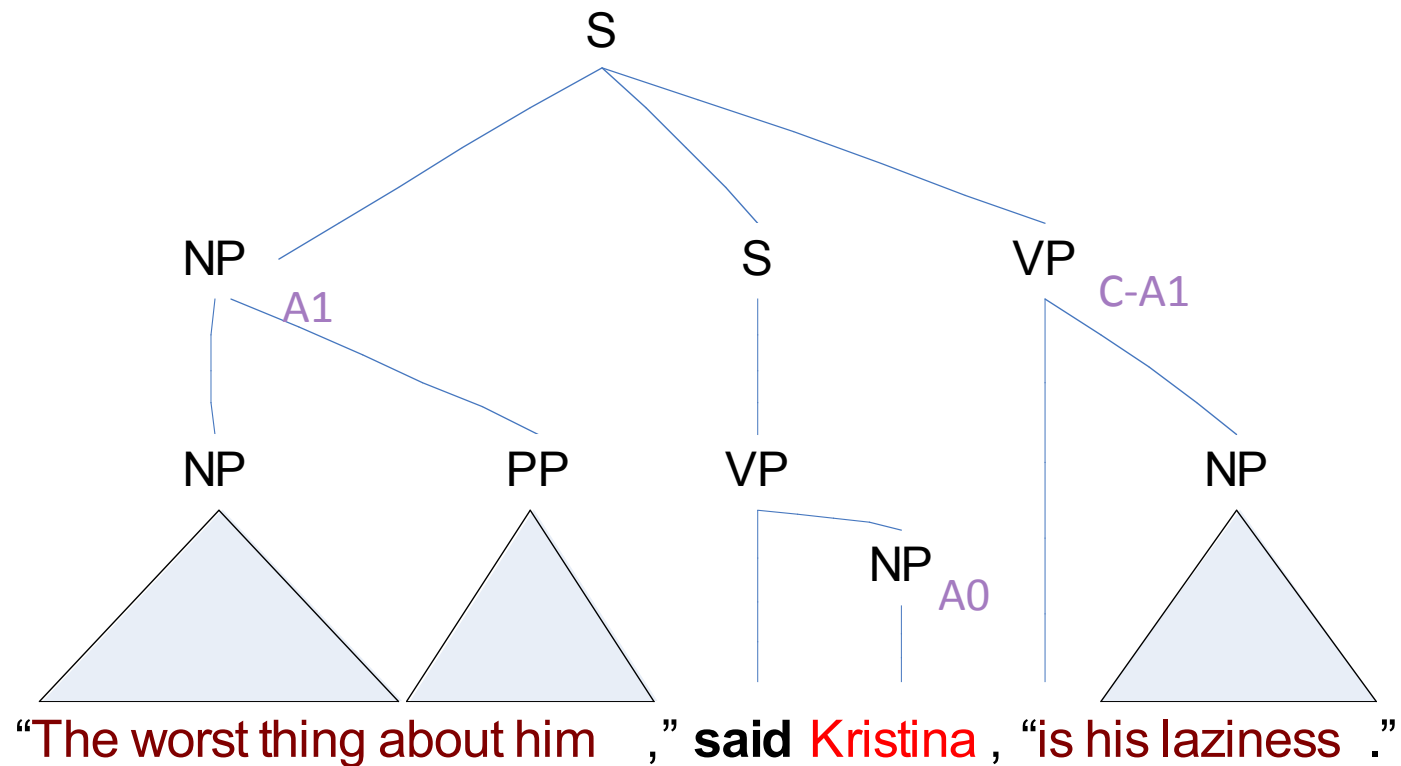 A sentence and
  a target verb

Proposition Bank (PropBank)
Add a Semantic Layer – Continued

[**A1** *The worst thing about him*] ***said*** [**A0** *Kristina*] [**C-A1** *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://verbs.colorado.edu/~mpalmer/projects/ace.html

- ## 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

Example

1. faces("the $1.4B robot spacecraft", "a six-year journey to explore Jupitor…")
2. 2. explore("the $1.4B robot spacecraft" "Jupiter …")

The	-	(A0*	(A0*
$1.4	-	*	*
billion	-	*	*
robot	-	*	*
spacecraft	-	*)	*)
faces	face	(V*)	*
a	-	(A1*	*
Six-year	-	*	*
journey	-	*	*
to	-	*	*
explore	expore	*	(V*)
Jupiter		*	(A1*
…		*)	*))

Example

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	-	(A0*)	(A0*)	*	*
is	-	*	*	*	*
accused	-	*	*	*	*
of	-	*	*	*	*
lying	lie	(V*)	*	*	*
under	-	*	*	*	*
oath	-	*	*	*	*
and	-	*	*	*	*
of	-	*	*	*	*
leaking	leak	*	(V*)	*	*
information	-	*	*	*	*
obtained	obtain	*	*	(V*)	*
from	-	*	*	*	*
a	-	*	*	*	(A1*
wiretap	-	*	*	*	*
he	-	*	*	*	(A0*)
supervised	supervise	*	*	*	(V*)

Information Extraction versus Semantic Role Labeling

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

Evaluation Measures

Correct: [A0 The queen] **broke** [A1 the window] [AM-TMP yesterday]

Guess: [A0 The queen] broke the [A1 window] [AM-LOC yesterday]

Correct	Guess
{The queen} →A0 {the window} →A1 {yesterday} ->AM-TMP all other → NONE	{The queen} →A0 {window} →A1 {yesterday} ->AM-LOC 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) {$tp=1,fp=1,fn=1$} $p=r=f=1/2$