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

Slide from Chris Manning
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.

Slide from Chris Manning
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
Common semantic roles

- **Goal:**
  - Bill ran to Copley Square.

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

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

- **Source:**
  - Bill took a pencil from the pile.

- **Instrument:**
  - Bill ate the burrito with a plastic spork.

- **Location:**
  - Bill sits under the tree 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?
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
+ Syntactic Variations (as trees)

Kristina hit Scott with a baseball yesterday

With a baseball, Kristina hit Scott yesterday
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)**
    - [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*

---

**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/fidb/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/ - 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/dipdf/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 - 7k - Cached - Similar pages

**NewsCentralAsia** - *Drugs in Afghanistan: Of carts and horses*
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
  - \( \{ \text{break}, \text{shatter}, \ldots \} \): 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.

<table>
<thead>
<tr>
<th>Alternation</th>
<th>Touch</th>
<th>Hit</th>
<th>Cut</th>
<th>Break</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conative</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Body-part possessor ascension</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Middle</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

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
Frames in Framenet
From Baker, Fillmore, Loew, 1998

Figure 1: A subframe can inherit elements and semantic from its parent

<table>
<thead>
<tr>
<th>Frame (TRANSPORTATION)</th>
</tr>
</thead>
<tbody>
<tr>
<td>frame_elements (MOVER(s), MEANS, PATH)</td>
</tr>
<tr>
<td>scene (MOVER(s) move along PATH by MEANS)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frame (DRIVING)</th>
</tr>
</thead>
<tbody>
<tr>
<td>inherit (TRANSPORTATION)</td>
</tr>
<tr>
<td>frame_elements (DRIVER (=MOVER), VEHICLE (=MEANS), RIDER(S) (=MOVER(S)), CARGO (=MOVER(S)))</td>
</tr>
<tr>
<td>scenes (DRIVER starts VEHICLE, DRIVER controls VEHICLE. DRIVER stops VEHICLE)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frame (RIDING_I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>inherit (TRANSPORTATION)</td>
</tr>
<tr>
<td>frame_elements (RIDER(s) (=MOVER(S)), VEHICLE (=MEANS))</td>
</tr>
<tr>
<td>scenes (RIDER enters VEHICLE, VEHICLE carries RIDER along PATH, RIDER leaves VEHICLE)</td>
</tr>
</tbody>
</table>
## Frame Element Groups

Figure 2: Examples of Frame Element Groups and Annotated Sentences

<table>
<thead>
<tr>
<th>FEG</th>
<th>Annotated Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>[D Kate] drove [P home] in a stupor.</td>
</tr>
<tr>
<td>D, V</td>
<td>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].</td>
</tr>
<tr>
<td>D, P</td>
<td>And that was why [D I] drove [P eastwards along Lake Geneva].</td>
</tr>
<tr>
<td>D, R, P</td>
<td>Now [D Van Checle] was driving [R his guest] [P back to the station].</td>
</tr>
<tr>
<td>D, V, P</td>
<td>[D Cumming] had a fascination with most forms of transport, driving [V his Rolls] at high speed [P around the streets of London].</td>
</tr>
<tr>
<td>D+R, P</td>
<td>[D We] drive [P home along miles of empty freeway].</td>
</tr>
<tr>
<td>V, P</td>
<td>Over the next 4 days, [V the Rolls Royces] will drive [P down to Plymouth], following the route of the railway.</td>
</tr>
</tbody>
</table>
**FrameNet** [Fillmore et al. 01]

**Frame: Hit_target**  
(hit, pick off, shoot)

<table>
<thead>
<tr>
<th>Core</th>
<th>Non-Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>Kristina</td>
</tr>
<tr>
<td>Target</td>
<td>hit</td>
</tr>
<tr>
<td>Instrument</td>
<td>Scott</td>
</tr>
<tr>
<td>Means</td>
<td>with a baseball</td>
</tr>
<tr>
<td>Place</td>
<td>Time</td>
</tr>
<tr>
<td>Purpose</td>
<td>yesterday</td>
</tr>
<tr>
<td>Manner</td>
<td>Subregion</td>
</tr>
</tbody>
</table>

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

**Frame elements (FEs):**  
The involved semantic roles
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-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] ...
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
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} \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

\[ \text{It looked to her like } A_0 \text{ he } \text{ deserved } A_1 \text{ this}. \]
Kristina hit Scott with a baseball yesterday.
The worst thing about him, said Kristina, “is his laziness.”

[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](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](http://www.lsi.upc.es/~srlconll/soft.html)
Example

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

<table>
<thead>
<tr>
<th></th>
<th>The</th>
<th>-</th>
<th>(A0*)</th>
<th>(A0*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1.4</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>billion</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>robot</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>spacecraft</td>
<td>-</td>
<td>*)</td>
<td>*)</td>
<td></td>
</tr>
<tr>
<td>faces</td>
<td>face</td>
<td>(V*)</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-</td>
<td>(A1*)</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>Six-year</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>journey</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>to</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>explore</td>
<td>explore</td>
<td>*</td>
<td>(V*)</td>
<td></td>
</tr>
<tr>
<td>Jupiter</td>
<td>*</td>
<td>(A1*)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td>*)</td>
<td>*)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 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")

<table>
<thead>
<tr>
<th></th>
<th>He</th>
<th>(A0*)</th>
<th>(A0*)</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>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>*</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>*</td>
<td>*</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>*</td>
<td>*</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>*</td>
<td>(A0*)</td>
</tr>
<tr>
<td>supervised</td>
<td>supervise</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>(V*)</td>
</tr>
<tr>
<td>Characteristic</td>
<td>IE</td>
<td>SRL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>--------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td>narrow</td>
<td>broad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth of semantics</td>
<td>shallow</td>
<td>shallow</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directly connected to application</td>
<td>sometimes</td>
<td>no</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Evaluation Measures

Correct: $[\text{A}_0 \text{The queen}]$ broke $[\text{A}_1 \text{the window}]$ $[\text{AM-TMP yesterday}]$

Guess: $[\text{A}_0 \text{The queen}]$ broke the $[\text{A}_1 \text{window}]$ $[\text{AM-LOC yesterday}]$

<table>
<thead>
<tr>
<th>Correct</th>
<th>Guess</th>
</tr>
</thead>
<tbody>
<tr>
<td>${\text{The queen}} \rightarrow \text{A}_0$</td>
<td>${\text{The queen}} \rightarrow \text{A}_0$</td>
</tr>
<tr>
<td>${\text{the window}} \rightarrow \text{A}_1$</td>
<td>${\text{window}} \rightarrow \text{A}_1$</td>
</tr>
<tr>
<td>${\text{yesterday}} \rightarrow \text{AM-TMP}$</td>
<td>${\text{yesterday}} \rightarrow \text{AM-LOC}$</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) $\text{acc} = .5$ (labeling of correctly identified phrases)
  - Core arguments (Precision, Recall, F-measure) $\{tp=1, fp=1, fn=1\}$ $p=r=f=1/2$