CS 4650/7650
Shallow Semantics

Jacob Eisenstein

October 28, 2013

The Roadmap

- **Compositional semantics**
  assemble the meaning of a sentence from its components
- **Shallow semantics**
  identify the key predicates and arguments in sentences
- **Lexical semantics**
  vector-space models for the meaning of individual words

Compositional semantics: pros and cons

- "Full" compositional semantics requires representations at least as expressive as first-order logic.
- Shallow semantics focuses on predicate-argument relations
- **Roles** are types of arguments.
  - Deep roles are predicate-specific.
  - Thematic roles are more general, but are hard to pin down.
  - Next we'll discuss semantic resources which address this issue.

Shallow semantics trades expressiveness for robustness and broader coverage.

Outline

Shallow semantics

Resources for shallow semantics

Semantic Role Labeling

SRL Today

Shallow semantics

- Consider these four sentences:
  - 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
  - Kristina hit Scott with a baseball yesterday

Shallow semantics

- Consider these four sentences:
  - 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
  - Kristina hit Scott with a baseball yesterday
- We don't need first-order logic to realize that these sentences are semantically identical.

Shallow semantics

- Consider these four sentences:
  - 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
  - Kristina hit Scott with a baseball yesterday
- We don't need first-order logic to realize that these sentences are semantically identical.
- Shallow semantics will suffice: the *roles* in each sentence are filled by the same text.
  - Hitter: Kristina
  - Person hit: Scott
  - Instrument of hitting: with a baseball
  - Time of hitting: yesterday

Deep roles

The event semantics representation for the sentence Scott was hit by Kristina yesterday (and all of the other examples) is:

$$\exists e, x, y\ \mathit{Hitting}(e) \wedge \mathit{Hitter}(e, \mathit{Kristina}) \wedge \mathit{PersonHit}(e, \mathit{Scott}) \\ \wedge\ \mathit{TimeOfHitting}(e, \mathit{Yesterday})$$

Deep roles

The event semantics representation for the sentence Scott was hit by Kristina yesterday (and all of the other examples) is:

$$\exists e, x, y \; Hitting(e) \wedge Hitter(e, Kristina) \wedge PersonHit(e, Scott) \\ \wedge \; TimeOfHitting(e, Yesterday)$$

- *Hitter*, *PersonHit*, and *TimeOfHitting* are roles.
- We use these specific roles because of the **predicate verb** hit.
- Roles that relate to a specific predicate are called "deep roles."

Thematic roles

Limitations of deep roles:

- Without knowing more about deep roles like *Hitter*, we cannot do much inference.
- Building classifiers for every role would be a lot of work.
- Consider Scott was paid by Kristina yesterday.
- Scott, Kristina and yesterday have similar thematic functions in each sentence.

Thematic roles

Limitations of deep roles:

- Without knowing more about deep roles like *Hitter*, we cannot do much inference.
- Building classifiers for every role would be a lot of work.
- Consider Scott was paid by Kristina yesterday.
- Scott, Kristina and yesterday have similar thematic functions in each sentence.

Thematic roles attempt to capture the similarity between *Payer* and *Hitter*, and between *PersonHit* and *PersonPaid*.

- Thematic roles date to Panini (7th-4th century BCE!)
- Modern formulation due to Fillmore (1968) and Gruber (1965)

Some typical thematic roles

AGENT	The volitional causer **The waiter** spilled the soup
EXPERIENCER	The experiencer The soup gave **all three of us** a headache.
FORCE	The non-volitional causer **The wind** blew my soup off the table.
THEME	The participant most directly affected The wind blew my **my soup** off the table.
RESULT	The end product The cook has prepared **a cold duck soup**.

Some typical thematic roles

CONTENT	The proposition or content of a propositional event The waiter assured me that **the soup is vegetarian**.
INSTRUMENT	An instrument used in an event It's hard to eat soup **with chopsticks**.
BENEFICIARY	The beneficiary The waiter brought **me** some soup.
SOURCE	Th origin of the object of a transfer event The stack of canned soup comes **from Pittsburgh**.
GOAL	The destination of the object of a transfer event He brought the bowl of soup **to our table**.

Case frames

- Different verbs take different thematic roles as arguments.

Case frames

- Different verbs take different thematic roles as arguments.
- The possible arguments for a verb is the **case frame** or **thematic grid**. For example, for break:
  - AGENT: Subject, THEME: Object
    John broke the window.
  - AGENT: Subject, THEME: Object, INSTRUMENT: PP (with)
    John broke the window with a rock.
  - INSTRUMENT: Subject, THEME: Object
    The rock broke the window.
  - THEME: Subject
    The window broke.

Case frames

- Different verbs take different thematic roles as arguments.
- The possible arguments for a verb is the **case frame** or **thematic grid**. For example, for break:
  - AGENT: Subject, THEME: Object
    John broke the window.
  - AGENT: Subject, THEME: Object, INSTRUMENT: PP (with)
    John broke the window with a rock.
  - INSTRUMENT: Subject, THEME: Object
    The rock broke the window.
  - THEME: Subject
    The window broke.
- When two verbs have similar case frames, this is a clue that they might be semantically related:
  (e.g., break, shatter, smash).

How many thematic roles?

- The purpose of thematic roles is to abstract above verb-specific roles.
- But it is usually possible to construct examples in which thematic roles are insufficiently specific.

How many thematic roles?

- The purpose of thematic roles is to abstract above verb-specific roles.
- But it is usually possible to construct examples in which thematic roles are insufficiently specific.
  - *Intermediary instruments* can act as subjects:
    1. The cook opened the jar with the new gadget.
    2. The new gadget opened the jar.
  - *Enabling instruments* cannot:
    1. Shelly ate the pizza with the fork.
    2. *The fork ate the pizza.

How many thematic roles?

- The purpose of thematic roles is to abstract above verb-specific roles.
- But it is usually possible to construct examples in which thematic roles are insufficiently specific.
  - *Intermediary instruments* can act as subjects:
    1. The cook opened the jar with the new gadget.
    2. The new gadget opened the jar.
  - *Enabling instruments* cannot:
    1. Shelly ate the pizza with the fork.
    2. *The fork ate the pizza.
- Thematic roles are bundles of semantic properties, but it's not clear how many properties are necessary.
  - AGENTS are usually animate, volitional, sentient, causal...
  - ...but any of these properties may be missing occasionally.

Outline

PropBank

In the Proposition Bank (**PropBank**), roles are verb-specific, with some sharing:

- Arg0: proto-agent (has agent-like properties)
- Arg1: proto-patient (has patient-like properties)
- Arg2... ArgN: verb-specific
- 13 universal adjunct-like arguments: temporal, manner, location, cause, negation, ...

PropBank

In the Proposition Bank (**PropBank**), roles are verb-specific, with some sharing:

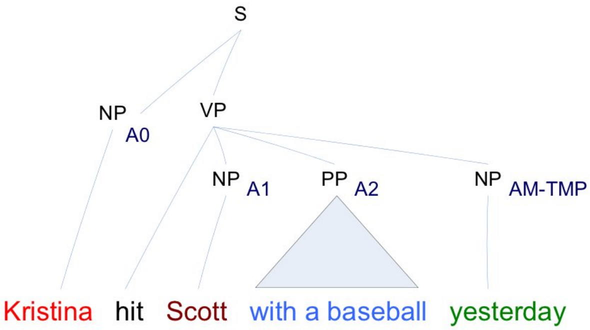
- Arg0: proto-agent (has agent-like properties)
- Arg1: proto-patient (has patient-like properties)
- Arg2... ArgN: verb-specific
- 13 universal adjunct-like arguments: temporal, manner, location, cause, negation, ...

PropBank contains two main resources:

- a set of labeled sentences, built on the Penn TreeBank
- a set of "Frame Files" describing each verbal predicate

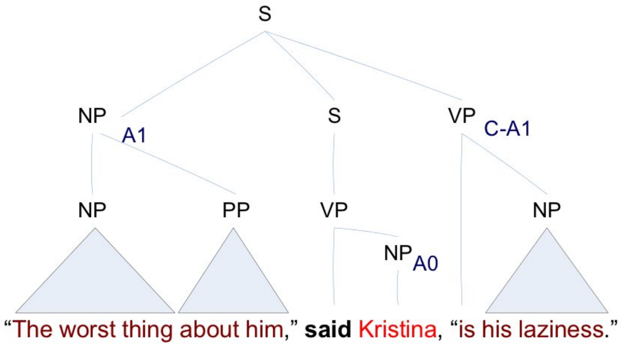
`http://verbs.colorado.edu/propbank/framesets-english/scratch-v.html`

Example PropBank annotations



$[_{A0}$ *Kristina*$]$ ***hit*** $[_{A1}$ *Scott*$]$ $[_{A2}$ *with a baseball*$]$ $[_{AM\text{-}TMP}$ *yesterday*$]$.

Example PropBank annotations



[$_{A1}$ *The worst thing about him*] ***said*** [$_{A0}$ *Kristina*] [$_{C\text{-}A1}$ *is his laziness*].

The PropBank corpus

- Last release: March 4, 2005
  - Verb Lexicon: 3,324 frame files
  - Annotation: 113,000 propositions
- PropBank has been used as the standard dataset for shared tasks on semantic role labeling (SRL)

The PropBank corpus

- Last release: March 4, 2005
  - Verb Lexicon: 3,324 frame files
  - Annotation: 113,000 propositions
- PropBank has been used as the standard dataset for shared tasks on semantic role labeling (SRL)
- Related corpora
  - Chinese PropBank
    `http://www.cis.upenn.edu/~chinese/cpb/`
  - NomBank: structure of noun phrases, e.g.
    $[_{A0}$ Her$]$ $[_{REL}$ gift$]$ of $[_{A1}$ a book$]$ $[_{A2}$ to John$]$

FrameNet

- Key idea: group related verbs (and nouns) into *frames*
  - [$_{A1}$ The price of bananas] increased [$_{A2}$ 5%].
  - [$_{A1}$ The price of bananas] rose [$_{A2}$ 5%].
  - There has been a [$_{A2}$ 5%] rise [$_{A1}$ in the price of bananas].
- First two sentences involve different verbs;
  second sentence conveys same semantics with a noun.
- Nonetheless, meaning is the same. FrameNet captures this.

FrameNet versus PropBank

FRAMENET ANNOTATION:

[$_{\text{Buyer}}$ Chuck] *bought* [$_{\text{Goods}}$ a car] [$_{\text{Seller}}$ from Jerry] [$_{\text{Payment}}$ for $1000].

[$_{\text{Seller}}$ Jerry] *sold* [$_{\text{Goods}}$ a car] [$_{\text{Buyer}}$ to Chuck] [$_{\text{Payment}}$ for $1000].

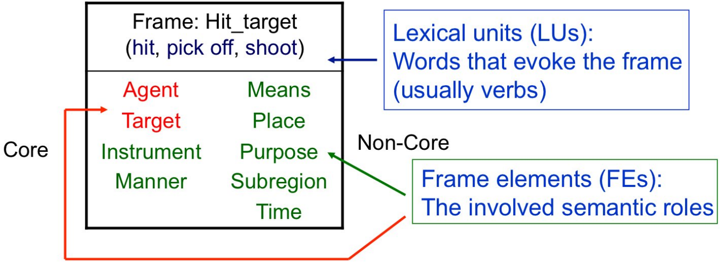
PROPBANK ANNOTATION:

[$_{\text{Arg0}}$ Chuck] *bought* [$_{\text{Arg1}}$ a car] [$_{\text{Arg2}}$ from Jerry] [$_{\text{Arg3}}$ for $1000].

[$_{\text{Arg0}}$ Jerry] *sold* [$_{\text{Arg1}}$ a car] [$_{\text{Arg2}}$ to Chuck] [$_{\text{Arg3}}$ for $1000].

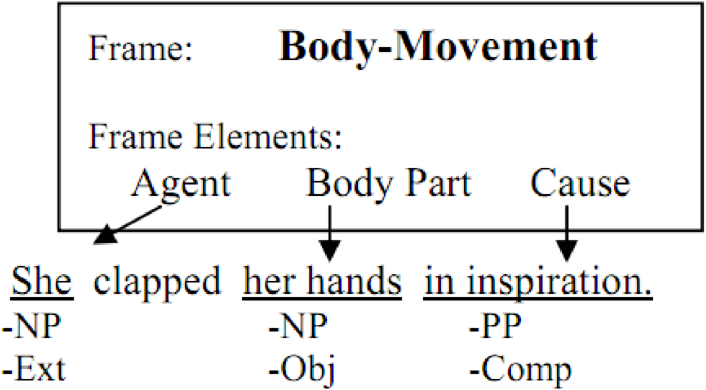
FrameNet

A Frame defines a set of *lexical units* and a set of *frame elements*:



[$_{\text{Agent}}$ *Kristina*] ***hit*** [$_{\text{Target}}$ *Scott*] [$_{\text{Instrument}}$ *with a baseball*] [$_{\text{Time}}$ *yesterday*].

FrameNet annotation



(figure from Fleischman et al, 2003)

`https://framenet.icsi.berkeley.edu/fndrupal/index.php?q=luIndex`

The FrameNet corpus

- https://framenet.icsi.berkeley.edu/fndrupal/about
- As of October 2013:
  - 1,164 semantic frames
  - 12,713 lexical units
  - 196,000 manually annotated sentences
  - still ongoing...
- Unlike PropBank,
  - not based on TreeBank parses
  - example sentences are chosen by hand

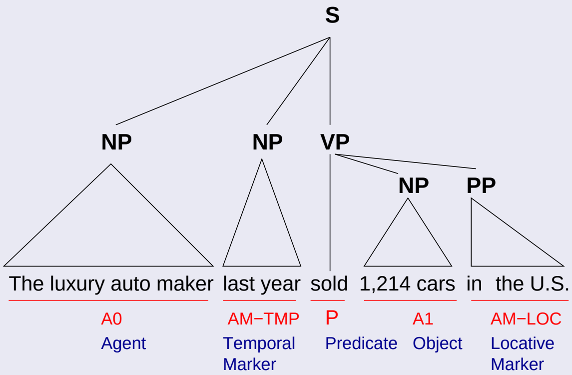
Outline

Semantic Role Labeling

- Semantic role labeling (SRL) is the task of assigning semantic labels to spans of text.
- Labels describe the role of the phrase with respect to the *predicate verb*.
- In practice, usually PropBank labels, e.g. Arg0

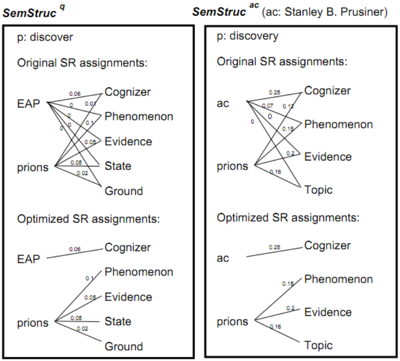
Semantic Role Labeling: The Problem

SRL $\stackrel{def}{=}$ detecting basic event structures such as *who* did *what* to *whom*, *when* and *where* [**IE** point of view]



Question answering

- Shen and Lapata (2007) use semantic roles to align questions against the content of factual sentences.
- Example:
  - Q: Who discovered prions?
  - S: 1997: Stanley B. Prusiner, United States, discovery of prions...



Subtasks

- **Identification**: determine which substrings are arguments
  - [$_{arg}$ Kristina] **hit** [$_{arg}$ Scott] [$_{arg}$ with a baseball] [$_{\varnothing}$ again]
  - In principle this is hard: lots of possible substrings.
  - In practice, parsing helps a lot. In PropBank,
    - 96% of arguments are a gold parse tree constituent
    - 90% of arguments are a (Charniak) parse tree constituent
    - Simple rules can recover the remaining arguments.

Subtasks

- **Identification**: determine which substrings are arguments
  - $[_{arg}$ Kristina$]$ **hit** $[_{arg}$ Scott$]$ $[_{arg}$ with a baseball$]$ $[_{\varnothing}$ again$]$
  - In principle this is hard: lots of possible substrings.
  - In practice, parsing helps a lot. In PropBank,
    - 96% of arguments are a gold parse tree constituent
    - 90% of arguments are a (Charniak) parse tree constituent
    - Simple rules can recover the remaining arguments.
- **Classification**: determine the label for each argument substring
  - $[_{A0}$ Kristina$]$ **hit** $[_{A1}$ Scott$]$ $[_{A2}$ with a baseball$]$ $[_{\varnothing}$ again$]$

Pipeline versus joint approaches

- **Pipeline**: first find arguments, then label them

$$\hat{a} = \arg\max_{a} P(a|\text{words}, \text{predicate})$$
$$\hat{y} = \arg\max_{y} P(y|\hat{a}, \text{words}, \text{predicate})$$

Pipeline versus joint approaches

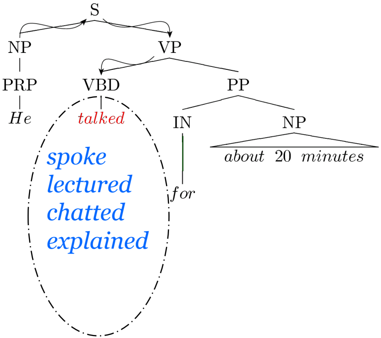
- **Pipeline**: first find arguments, then label them

$$\hat{a} = \arg\max_a P(a|\text{words}, \text{predicate})$$
$$\hat{y} = \arg\max_y P(y|\hat{a}, \text{words}, \text{predicate})$$

- **Joint**: compute arguments and labels jointly

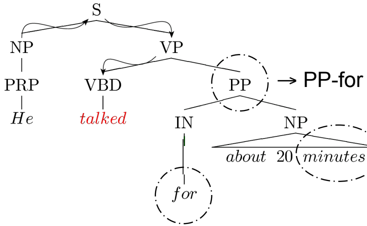
$$\begin{aligned}\hat{y} &= \arg\max_y P(y|\text{words}, \text{predicate}) \\ &= \arg\max_y \sum_a P(y, a|\text{words}, \text{predicate}) \\ &= \arg\max_y \sum_a P(y|a, \text{words}, \text{predicate}) P(a|\text{words}, \text{predicate})\end{aligned}$$

Predicate cluster, automatic or WordNet

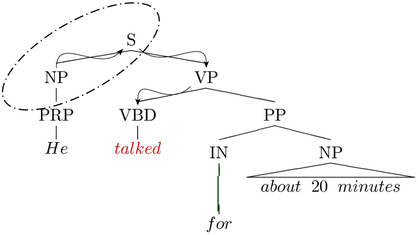


Slide from Sameer Pradhan

Noun Head and POS of PP

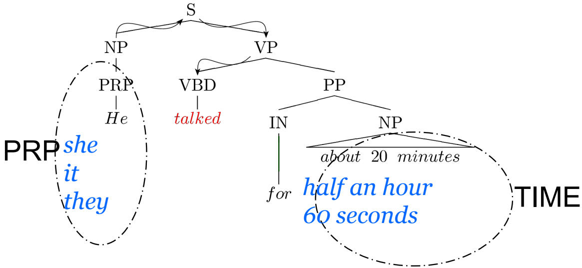


Partial Path

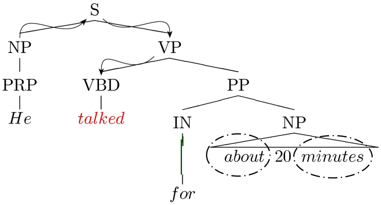


Named Entities and Head Word POS

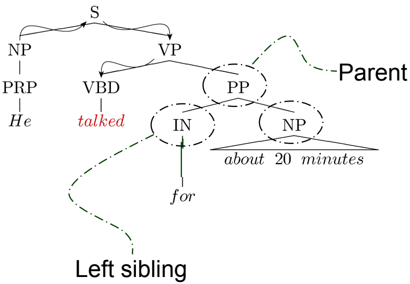
[Surdeanu et al., 2003]



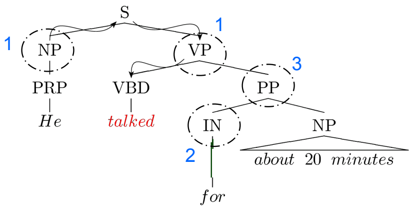
First and Last Word and POS



Parent and Sibling features



Constituent tree distance



Combining local and global scoring

- Individual labels may each be good, but they may not fit together well.
- Global scoring checks the **overall** labeling. Some approaches:
  - Local scoring, then **re-rank** (Gildea and Jurafsky 2002, Toutanova et al 2005)

Combining local and global scoring

- Individual labels may each be good, but they may not fit together well.
- Global scoring checks the **overall** labeling. Some approaches:
  - Local scoring, then **re-rank** (Gildea and Jurafsky 2002, Toutanova et al 2005)
  - **Joint probability model** with some independence assumptions (e.g., TreeCRF of Blunsom et al 2004)

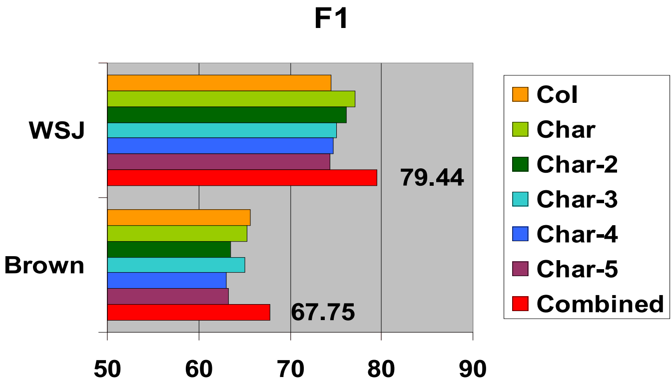
Combining local and global scoring

- Individual labels may each be good, but they may not fit together well.
- Global scoring checks the **overall** labeling. Some approaches:
  - Local scoring, then **re-rank** (Gildea and Jurafsky 2002, Toutanova et al 2005)
  - **Joint probability model** with some independence assumptions (e.g., TreeCRF of Blunsom et al 2004)
  - Do exact search for best local model satisfying **global constraints** (Punyakanok et al, 2004)

Global constraints for SRL

- Many of the global criteria can be viewed as constraints:
  - Arguments may not overlap.
  - No argument type may appear twice.
  - Arguments do not overlap the predicate.
- Only some constraints can be built into a dynamic program.
- Instead, solve as a constrained optimization problem using Integer Linear Programming (ILP). (see notes)

Generalized Inference – ILP (Koomen et al., 2005; Punyakanok et al., 2008)



- Inference with many parsers improves results ~ 2.6 F_1 points
- Best results at CoNLL-2005 shared task (Carreras & Màrquez, 2005)

Outline

Open Issues for SRL

SRL degrades badly when moved to new domains.

- Moving from WSJ to Brown test corpus causes F-measure to decrease from 80% to 70%.
- The decline is mainly due to role classification, not argument identification (Pradhan et al, 2008).
- One explanation:
  SRL is high in the "food chain" — it consumes the output of many other NLP systems. If POS tagging, parsing, or WSD get worse, then SRL will too.
- Another explanation:
  Lexical semantics is more domain-specific than syntax?

Open Issues for SRL

SRL depends on parsing, and may cascade parsing errors.
Can we do parsing and SRL jointly?

- K-best parses (Sutton and McCallum 2005)
- Parse sampling (Finkel et al 2006)
- Synchronouns dependency parsing for syntax and semantics (Gesmundo et al 2009)

Open Issues for SRL

SRL depends on parsing, and may cascade parsing errors.
Can we do parsing and SRL jointly?

- K-best parses (Sutton and McCallum 2005)
- Parse sampling (Finkel et al 2006)
- Synchronouns dependency parsing for syntax and semantics (Gesmundo et al 2009)

Supervised SRL requires expensive resources.
What about unsupervised learning?

- split-merge clustering to identify verb alternations (Lang and Lapata 2010, 2011)
- non-parametric Bayesian model of predicates and roles (Titov and Klementiev 2012)

Resources for SRL

- Online demo:
  http://cogcomp.cs.illinois.edu/demo/srl/
- Corpora:
  - PropBank sold by Linguistic Data Consortium ($500); you will also need Penn TreeBank
  - FrameNet available on the web
    https://framenet.icsi.berkeley.edu/fndrupal/
- SRL systems
  - http://cogcomp.cs.illinois.edu/page/software_view/SRL
  - http://cemantix.org/assert.html
  - http://www.coli.uni-saarland.de/projects/salsa/shal/

Recap

- Shallow semantics represents meaning through **predicate-argument** structures.
  - **Thematic roles** are argument types that shared across many predicates.
  - PropBank uses some proto-roles and some verb-specific roles.
  - FrameNet groups verbs into semantic frames
- Semantic Role Labeling is the task of identifying and labeling the semantic arguments to each predicate in a sentence.
- High-quality SRL requires global inference, which can be performed using constrained optimization (ILP).