Frame Semantics
Luke Zettlemoyer

Slides adapted from Yejin Choi, Martha Palmer, Chris Manning, Ray Mooney, Lluis Marquez, Luheng He
Frames

- **Theory:**
  - Frame Semantics (Fillmore 1968)

- **Resources:**
  - VerbNet (Kipper et al., 2000)
  - FrameNet (Fillmore et al., 2004)
  - PropBank (Palmer et al., 2005)
  - NomBank

- **Statistical Models:**
  - Task: Semantic Role Labeling (SRL)
  - Deep SRL

"Case for Case"
If you got a billion dollars to spend on a huge research project that you get to lead, what would you like to do?

I'd use the billion dollars to build a **NASA-size program** focusing on natural language processing (NLP), in all of its glory (**semantics**, **pragmatics**, etc).

Intellectually I think that NLP is fascinating, allowing us to focus on **highly-structured inference problems**, on issues that go to the core of "**what is thought**" but remain eminently practical, and on a technology that surely would make the world a better place.
Although current deep learning research tends to claim to encompass NLP, I'm (1) much less convinced about the strength of the results, compared to the results in, say, vision; (2) much less convinced in the case of NLP than, say, vision, the way to go is to couple huge amounts of data with black-box learning architectures.

I'd invest in some of the human-intensive labeling processes that one sees in projects like FrameNet and (gasp) projects like Cyc. I'd do so in the context of a full merger of "data" and "knowledge", where the representations used by the humans can be connected to data and the representations used by the learning systems are directly tied to linguistic structure. I'd do so in the context of clear concern with the usage of language (e.g., causal reasoning).
Frames

Theory:
- Frame Semantics (Fillmore 1968)

Resources:
- VerbNet (Kipper et al., 2000)
- FrameNet (Fillmore et al., 2004)
- PropBank (Palmer et al., 2005)
- NomBank

Statistical Models:
- Task: Semantic Role Labeling (SRL)
- Deep SRL

“Case for Case”
Frame Semantics

- **Frame**: Semantic frames are schematic representations of situations involving various participants, propositions, and other conceptual roles.
- **Frame Elements (FEs)** include events, states, relations, and entities.

- **Frame**: “The case for case” (Fillmore 1968)
  - 8k citations in Google Scholar.

- **Script**: knowledge about situations like eating in a restaurant.
  - “Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures” (Schank & Abelson 1977)

- **Political Framings**: George Lakoff’s recent writings on the framing of political discourse.
### Capturing Generalizations over Related Predicates & Arguments

<table>
<thead>
<tr>
<th>verb</th>
<th>BUYER</th>
<th>GOODS</th>
<th>SELLER</th>
<th>MONEY</th>
<th>PLACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy</td>
<td>subject</td>
<td>object</td>
<td>from</td>
<td>for</td>
<td>at</td>
</tr>
<tr>
<td>Sell</td>
<td>to</td>
<td>object</td>
<td>subject</td>
<td>for</td>
<td>at</td>
</tr>
<tr>
<td>Cost</td>
<td>Ind. object</td>
<td>subject</td>
<td>--</td>
<td>object</td>
<td>at</td>
</tr>
<tr>
<td>Spend</td>
<td>subject</td>
<td>on</td>
<td>--</td>
<td>object</td>
<td>at</td>
</tr>
</tbody>
</table>
Case Grammar -> Frames

- Valency: Predicates have arguments (optional & required)
  - Example: “give” requires 3 arguments:
    - Agent (A), Object (O), and Beneficiary (B)
    - Jones (A) gave money (O) to the school (B)

- Frames:
  - commercial transaction frame: Buy/Sell/Pay/Spend
  - Save <good thing> from <bad situation>
  - Risk <valued object> for
    <situation>|<purpose>|<beneficiary>|<motivation>

- Collocations & Typical predicate argument relations
  - Save whales from extinction (not vice versa)
  - Ready to risk everything for what he believes

- Representation Challenges: What matters for practical NLP?

Slide from Ken Church (at Fillmore tribute workshop)
Thematic (Semantic) Roles

- **AGENT** - the volitional causer of an event
  - The waiter spilled the soup
- **EXPERIENCER** - the experiencer of an event
  - John has a headache
- **FORCE** - the non-volitional causer of an event
  - The wind blows debris from the mall into our yards.
- **THEME** - the participant most directly affected by an event
  - Only after Benjamin Franklin broke the ice ...
- **RESULT** - the end product of an event
  - The French government has built a regulation-size baseball diamond ...
Thematic (Semantic) Roles

- **INSTRUMENT** - an instrument used in an event
  - He turned to poaching catfish, stunning them with a shocking device...

- **BENEFICIARY** - the beneficiary of an event
  - Whenever Ann makes hotel reservations for her boss...

- **SOURCE** - the origin of the object of a transfer event
  - I flew in from Boston

- **GOAL** - the destination of an object of a transfer event
  - I drove to Portland

- Can we read semantic roles off from PCFG or dependency parse trees?
Semantic roles $\neq$ Grammatical roles

- **Agent** – the volitional causer of an event
  - usually “subject”, sometimes “prepositional argument”, ...
- **Theme** – the participant directly affected by an event
  - usually “object”, sometimes “subject”, ...
- **Instrument** – an instrument (method) used in an event
  - usually prepositional phrase, but can also be a “subject”

- John broke the window.
- John broke the window with a rock.
- The rock broke the window.
- The window broke.
- The window was broken by John.
Ergative Verbs

- Ergative verbs
  - subject when intransitive = direct object when transitive.
  - "it broke the window" (transitive)
  - "the window broke" (intransitive).
- Most verbs in English are not ergative (the subject role does not change whether transitive or not)
  - "He ate the soup" (transitive)
  - "He ate" (intransitive)
- Ergative verbs generally describe some sort of “changes” of states:
  - Verbs suggesting a change of state — break, burst, form, heal, melt, tear, transform
  - Verbs of cooking — bake, boil, cook, fry
  - Verbs of movement — move, shake, sweep, turn, walk
  - Verbs involving vehicles — drive, fly, reverse, run, sail
FrameNet
Frames

Theory:
- Frame Semantics (Fillmore 1968)

Resources:
- VerbNet (Kipper et al., 2000)
- FrameNet (Fillmore et al., 2004)
- PropBank (Palmer et al., 2005)
- NomBank

Statistical Models:
- Task: Semantic Role Labeling (SRL)

“Case for Case”
Words in “change_position_on_a_scale” frame:

<table>
<thead>
<tr>
<th>VERBS:</th>
<th>dwindle</th>
<th>move</th>
<th>soar</th>
<th>escalation</th>
<th>shift</th>
<th>advance</th>
<th>edge</th>
<th>mushroom</th>
<th>swell</th>
<th>explosion</th>
<th>tumble</th>
</tr>
</thead>
<tbody>
<tr>
<td>climb</td>
<td>explode</td>
<td>plummet</td>
<td>swing</td>
<td>fall</td>
<td></td>
<td>decrease</td>
<td>fall</td>
<td>reach</td>
<td>triple</td>
<td>fluctuation</td>
<td>tumble</td>
</tr>
<tr>
<td>decrease</td>
<td>fluctuate</td>
<td>rise</td>
<td>tumble</td>
<td>gain</td>
<td>increasingly</td>
<td>dip</td>
<td>grow</td>
<td>shift</td>
<td>hike</td>
<td></td>
<td></td>
</tr>
<tr>
<td>drop</td>
<td>jump</td>
<td>slide</td>
<td>decrease</td>
<td>rise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Frame := the set of words sharing a similar predicate-argument relations
- Predicate can be a verb, noun, adjective, adverb
- The same word with multiple senses can belong to multiple frames
Roles in “\textit{change\_position\_on\_a\_scale}” frame

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ATTRIBUTE</strong></td>
<td>The ATTRIBUTE is a scalar property that the ITEM possesses.</td>
</tr>
<tr>
<td><strong>DIFFERENCE</strong></td>
<td>The distance by which an ITEM changes its position on the scale.</td>
</tr>
<tr>
<td><strong>FINAL_STATE</strong></td>
<td>A description that presents the ITEM’s state after the change in the ATTRIBUTE’s value as an independent predication.</td>
</tr>
<tr>
<td><strong>FINAL_VALUE</strong></td>
<td>The position on the scale where the ITEM ends up.</td>
</tr>
<tr>
<td><strong>INITIAL_STATE</strong></td>
<td>A description that presents the ITEM’s state before the change in the ATTRIBUTE’s value as an independent predication.</td>
</tr>
<tr>
<td><strong>INITIAL_VALUE</strong></td>
<td>The initial position on the scale from which the ITEM moves away.</td>
</tr>
<tr>
<td><strong>ITEM</strong></td>
<td>The entity that has a position on the scale.</td>
</tr>
<tr>
<td><strong>VALUE_RANGE</strong></td>
<td>A portion of the scale, typically identified by its end points, along which the values of the ATTRIBUTE fluctuate.</td>
</tr>
</tbody>
</table>

**Some Non-Core Roles**

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DURATION</strong></td>
<td>The length of time over which the change takes place.</td>
</tr>
<tr>
<td><strong>SPEED</strong></td>
<td>The rate of change of the VALUE.</td>
</tr>
<tr>
<td><strong>GROUP</strong></td>
<td>The GROUP in which an ITEM changes the value of an ATTRIBUTE in a specified way.</td>
</tr>
</tbody>
</table>
Example

- [Oil] rose [in price] [by 2%].

- [It] has increased [to having them 1 day a month].

- [Microsoft shares] fell [to 7 5/8].

- [cancer incidence] fell [by 50%] [among men].

- a steady increase [from 9.5] [to 14.3] [in dividends].

- a [5%] [dividend] increase…
Find “Item” roles?

- [Oil] rose [in price] [by 2%].
- [It] has increased [to having them] [1 day a month].
- [Microsoft shares] fell [to 7 5/8].
- [cancer incidence] fell [by 50%] [among men].
- a steady increase [from 9.5] [to 14.3] [in dividends].
- a [5%] [dividend] increase…
Find “Difference” & “Final_Value” roles?

- [Oil] rose [in price] [by 2%].
- [It] has increased [to having them] [1 day a month].
- [Microsoft shares] fell [to 7 5/8].
- [cancer incidence] fell [by 50%] [among men].
- a steady increase [from 9.5] [to 14.3] [in dividends].
- a [5%] [dividend] increase…
FrameNet (2004)

- Project at UC Berkeley led by Chuck Fillmore for developing a database of frames, general semantic concepts with an associated set of roles.
- Roles are specific to frames, which are “invoked” by the predicate, which can be a verb, noun, adjective, adverb
  - **JUDGEMENT frame**
    - Invoked by: V: blame, praise, admire; N: fault, admiration
    - Roles: JUDGE, EVALUTEE, and REASON
- Specific frames chosen, and then sentences that employed these frames selected from the British National Corpus and annotated by linguists for semantic roles.
- Initial version: 67 frames, 49,013 sentences, 99,232 role fillers
PropBank
(proposition bank)
PropBank := proposition bank (2005)

- Project at Colorado led by Martha Palmer to add semantic roles to the Penn treebank.
- Proposition := verb + a set of roles
- Annotated over 1M words of Wall Street Journal text with existing gold-standard parse trees.
- Statistics:
  - 43,594 sentences       99,265 propositions
  - 3,324 unique verbs     262,281 role assignments
PropBank argument numbering

- Numbered roles, rather than named roles.
  - Arg0, Arg1, Arg2, Arg3, ...

- Different numbering scheme for each verb sense.

- The general pattern of numbering is as follows.

- Arg0 = “Proto-Agent” (agent)
- Arg1 = “Proto-Patient” (direct object / theme / patient)
- Arg2 = indirect object (benefactive / instrument / attribute / end state)
- Arg3 = start point (benefactive / instrument / attribute)
- Arg4 = end point
Different “frameset” for each verb sense

- Mary left the room.
- Mary left her daughter-in-law her pearls in her will.

Frameset **leave.01** "move away from":
Arg0: entity leaving
Arg1: place left

Frameset **leave.02** "give":
Arg0: giver
Arg1: thing given
Arg2: beneficiary
Semantic Role Labeling
Semantic Role Labeling (Task)

- Shallow meaning representation beyond syntactic parse trees
- Question Answering
  - “Who” questions usually use Agents
  - “What” question usually use Patients
  - “How” and “with what” questions usually use Instruments
  - “Where” questions frequently use Sources and Destinations.
  - “For whom” questions usually use Beneficiaries
  - “To whom” questions usually use Destinations
- Machine Translation Generation
  - Semantic roles are usually expressed using particular, distinct syntactic constructions in different languages.
- Summarization, Information Extraction
SRL $^\text{def}$ detecting basic event structures such as *who did what to whom, when and where* [IE point of view]

Example from Lluis Marquez
SRL \(\overset{def}{=}\)

identify the *arguments* of a given verb and assign them semantic labels describing the *roles* they play in the predicate (i.e., identify predicate argument structures) [CL point of view]
Linguistic nature of the problem

- Argument identification is strongly related to syntax

  ![Syntax Tree Example]

  - NP: The luxury auto maker
  - NP: sold
  - VP: 1,214 cars
  - NP: in
  - NP: the U.S.
  - PP: last year
  - A0: Agent
  - AM-TMP: Temporal Marker
  - P: Predicate
  - A1: Object
  - AM-LOC: Locative Marker

- Role labeling is a semantic task
  - e.g., selectional preferences should play an important role
SRL as Parse Node Classification

- Assume that a syntactic parse is available
- Treat problem as classifying parse-tree nodes.
- Can use any machine-learning classification method.
- Critical issue is engineering the right set of features for the classifier to use.

**Color Code:**
- not-a-role
- agent
- patient
- source
- destination
- instrument
- beneficiary
Deep Semantic Role Labeling
SRL Systems

**Pipeline Systems**

- sentence, predicate
- syntactic features
- argument id.
- candidate argument spans
- labeling
- labeled arguments
- ILP/DP
- prediction

Punyakanok et al., 2008
Täckström et al., 2015
FitzGerald et al., 2015

**End-to-end Systems**

- sentence, predicate
- context window features
- Deep BiLSTM + CRF layer
- BIO sequence
- Viterbi
- prediction

Collobert et al., 2011
Zhou and Xu, 2015
Wang et al., 2015

**Most Recent Work**

- sentence, predicate
- Deep BiLSTM
- BIO sequence
- Hard constraints
- prediction

He et al. 2017, 2018
SRL as BIO Tagging Problem

Input (sentence and predicate):

BIO output: (Begin, Inside, Outside)

Final SRL output:
(1) Deep BiLSTM tagger

(2) Highway connections

(3) Variational dropout

(4) Viterbi decoding with hard constraints

(0) Embeddings / predicate ID

(1) Deep BiLSTM tagger

(2) Highway connections

(3) Variational dropout

(4) Viterbi decoding with hard constraints
Model - (2) Highway Connections

Trend: Deeper models for higher accuracy

Grammar as a Foreign Language (Vinyals et al., 2014): 3 layers
End-to-end Semantic Role Labeling (Zhou and Xu, 2015): 8 layers
Google’s Neural Machine Translation (GNMT, Wu et al., 2016): 8 layers
Deep Semantic Role Labeling (He et al 2017): 8 layers
Deep Residual Learning for Image Recognition (He et al, 2016): 152 layers
the cats love hats

increase expressive power

harder to back-propagate
Model - (2) Highway Connections

\[ h_t \]  
\[ x_t \]  
\[ r_t \circ h_t + (1 - r_t) \circ x_t \]  
\[ r_t = \sigma(f(h_{t-1}, x_t)) \]

References:
Deep Residual Networks, Kaiming He, ICML 2016 Tutorial
Training Very Deep Networks, Srivastava et al., 2015
Traditionally, dropout masks are only applied to vertical connections.

Applying dropout to recurrent connections causes too much noise amplification.

**Variational dropout**: Reuse the same dropout mask for each timestep. 
Gal and Ghahramani, 2016
Model - (4) Viterbi Decoding with Hard Constraints

Viterbi decoding

<table>
<thead>
<tr>
<th></th>
<th>B-ARG0</th>
<th>I-ARG0</th>
<th>B-ARG1</th>
<th>I-ARG1</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-ARG0</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-ARG0</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-ARG1</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-ARG1</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BIO inconsistency

<table>
<thead>
<tr>
<th></th>
<th>B-ARG0</th>
<th>I-ARG0</th>
<th>B-ARG1</th>
<th>I-ARG1</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-ARG0</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-ARG0</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-ARG1</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-ARG1</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Greedy Output

argmax

Softmax

BiLSTM layers ...

the [ ]
cats [ ]
love [V]
hats [ ]

Model - (4) Viterbi Decoding with Hard Constraints
Other Implementation Details …

- 8 layer BiLSTMs with 300D hidden layers.
- 100D GloVe embeddings, updated during training.
- **Orthonormal initialization** for LSTM weight matrices (Saxe et al., 2013)
- 5 model ensemble with **product-of-experts** (Hinton 2002)
- Trained for 500 epochs.
CoNLL 2005 Results

WSJ Test

Brown (out-domain) Test

Ours*

Zhou

Täckström

Punyakanok*

BiLSTM models

Pipeline models

*: Ensemble models
Ablations on Number of Layers
(2, 4, 6 and 8)

Shallow models benefit more from constrained decoding.

Performance increases as model goes deeper. Biggest jump from 2 to 4 layer.
Ablations (single model)

- Full model
- No highway
- No orthonormal init.
- No dropout

Without dropout, model overfits at ~300 epochs.

Without initialization, the deep model learns very slowly.