Semantic Role Labeling

Introduction
Semantic Role Labeling

Who | did what to whom | at where?

The police officer detained the suspect at the scene of the crime

Agent | Predicate | Theme | Location

Who did what to whom at where?

Question & answer systems

Who did what to whom at where?

The police officer detained the suspect at the scene of the crime

Agent | Predicate | Theme | Location
Can we figure out that these have the same meaning?

XYZ corporation bought the stock.
They sold the stock to XYZ corporation.
The stock was bought by XYZ corporation.
The purchase of the stock by XYZ corporation...
The stock purchase by XYZ corporation...
A Shallow Semantic Representation: Semantic Roles

Predicates (bought, sold, purchase) represent an event. Semantic roles express the abstract role that arguments of a predicate can take in the event.

More specific

buyer  agent  proto-agent

More general
Semantic Role Labeling

Semantic Roles
Getting to semantic roles

Neo-Davidsonian event representation:

Sasha broke the window
Pat opened the door

Subjects of break and open: Breaker and Opener

Deep roles specific to each event (breaking, opening)

Hard to reason about them for NLU applications like QA
Thematic roles

- **Breaker** and **Opener** have something in common!
  - Volitional actors
  - Often animate
  - Direct causal responsibility for their events
- Thematic roles are a way to capture this semantic commonality between **Breakers** and **Eaters**.
- They are both **AGENTS**.
- The **BrokenThing** and **OpenedThing**, are **THEMES**.
  - prototypically inanimate objects affected in some way by the action
Thematic roles

• One of the oldest linguistic models
  • Indian grammarian Panini between the 7th and 4th centuries BCE

• Modern formulation from Fillmore (1966,1968), Gruber (1965)
  • Fillmore influenced by Lucien Tesnière’s (1959) Éléments de Syntaxe Structurale, the book that introduced dependency grammar
  • Fillmore first referred to roles as actants (Fillmore, 1966) but switched to the term case
Thematic roles

- A typical set:

<table>
<thead>
<tr>
<th>Thematic Role</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGENT</td>
<td>The volitional causer of an event</td>
<td><em>The waiter</em> spilled the soup.</td>
</tr>
<tr>
<td>EXPERIENCER</td>
<td>The experiencer of an event</td>
<td><em>John</em> has a headache.</td>
</tr>
<tr>
<td>FORCE</td>
<td>The non-volitional causer of the event</td>
<td><em>The wind</em> blows debris from the mall into our yards.</td>
</tr>
<tr>
<td>THEME</td>
<td>The participant most directly affected by an event</td>
<td>Only after Benjamin Franklin broke <em>the ice</em>...</td>
</tr>
<tr>
<td>RESULT</td>
<td>The end product of an event</td>
<td>The city built a <em>regulation-size baseball diamond</em>...</td>
</tr>
<tr>
<td>CONTENT</td>
<td>The proposition or content of a propositional event</td>
<td>Mona asked “<em>You met Mary Ann at a supermarket?</em>”</td>
</tr>
<tr>
<td>INSTRUMENT</td>
<td>An instrument used in an event</td>
<td>He poached catfish, stunning them with a <em>shocking device</em>...</td>
</tr>
<tr>
<td>BENEFICIARY</td>
<td>The beneficiary of an event</td>
<td>Whenever Ann Callahan makes hotel reservations for <em>her boss</em>...</td>
</tr>
<tr>
<td>SOURCE</td>
<td>The origin of the object of a transfer event</td>
<td>I flew in <em>from Boston</em>.</td>
</tr>
<tr>
<td>GOAL</td>
<td>The destination of an object of a transfer event</td>
<td>I drove to <em>Portland</em>.</td>
</tr>
</tbody>
</table>
Diathesis Alternations

Example usages of “break”

\[\begin{align*}
  \text{John} & \quad \textit{broke the window.} & \text{AGENT} & \quad \text{THEME} \\
  \text{John} & \quad \textit{broke the window with a rock.} & \text{AGENT} & \quad \text{THEME} & \quad \text{INSTRUMENT} \\
  \text{The rock} & \quad \textit{broke the window.} & \text{INSTRUMENT} & \quad \text{THEME} \\
  \text{The window} & \quad \textit{broke.} & \text{THEME} \\
  \text{The window was broken by John.} & \text{THEME} & \quad \text{AGENT}
\end{align*}\]

\[\text{Break:}\]
\[\text{AGENT, THEME, INSTRUMENT.}\]

Some realizations:

\[\begin{align*}
  & \text{AGENT/Subject, THEME/Object} \\
  & \text{AGENT/Subject, THEME/Object, INSTRUMENT/PP with} \\
  & \text{INSTRUMENT/Subject, THEME/Object} \\
  & \text{THEME/Subject}
\end{align*}\]
Diathesis alternations (or verb alternation)

Doris gave the book to Cary.  
AGENT THEME GOAL

Doris gave Cary the book.  
AGENT GOAL THEME

**Dative alternation**: particular semantic classes of verbs, “verbs of future having” (advance, allocate, offer, owe), “send verbs” (forward, hand, mail), “verbs of throwing” (kick, pass, throw), etc.

**Levin (1993)**: 47 semantic classes (“Levin classes”) for 3100 English verbs and alternations. In online resource VerbNet.

---

*Break*: AGENT, INSTRUMENT, or THEME as subject

*Give*: THEME and GOAL in either order
Problems with Thematic Roles

Hard to create standard set of roles or formally define them. Often roles need to be fragmented to be defined.

Levin and Rappaport Hovav (2015): two kinds of INSTRUMENTS

* intermediary instruments that can appear as subjects
  The cook opened the jar with the new gadget.
  The new gadget opened the jar.

* enabling instruments that cannot
  Shelly ate the sliced banana with a fork.
  *The fork ate the sliced banana.
Alternatives to thematic roles

1. **Fewer roles**: generalized semantic roles, defined as prototypes (Dowty 1991)
   
   PROTO-AGENT
   PROTO-PATIENT
   PropBank

2. **More roles**: Define roles specific to a group of predicates
   
   FrameNet
Semantic Role Labeling

The Proposition Bank (PropBank)
PropBank Roles

Proto-Agent

- Volitional involvement in event or state
- Sentience (and/or perception)
- Causes an event or change of state in another participant
- Movement (relative to position of another participant)

Proto-Patient

- Undergoes change of state
- Causally affected by another participant
- Stationary relative to movement of another participant

Following Dowty 1991
PropBank Roles

- Following Dowty 1991
  - Role definitions determined verb by verb, with respect to the other roles
  - Semantic roles in PropBank are thus verb-sense specific.
- Each verb sense has numbered argument: Arg0, Arg1, Arg2,
  Arg0: PROTO-AGENT
  Arg1: PROTO-PATIENT
  Arg2: usually: benefactive, instrument, attribute, or end state
  Arg3: usually: start point, benefactive, instrument, or attribute
  Arg4 the end point

(Arg2-Arg5 are not really that consistent, causes a problem for labeling)
PropBank Frame Files

agree.01
Arg0: Agreer
Arg1: Proposition
Arg2: Other entity agreeing

Ex1: [Arg0 The group] agreed [Arg1 it wouldn’t make an offer].
Ex2: [ArgM-TMP Usually] [Arg0 John] agrees [Arg2 with Mary] [Arg1 on everything].

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 $25 million] [Arg3 from $27 million].
Ex2: [Arg1 The average junk bond] fell [Arg2 by 4.2%].
Advantage of a ProbBank Labeling

`increase.01 “go up incrementally”`

Arg0: causer of increase
Arg1: thing increasing
Arg2: amount increased by, EXT, or MNR
Arg3: start point
Arg4: end point

This would allow us to see the commonalities in these 3 sentences:

[Arg0 Big Fruit Co.] increased [Arg1 the price of bananas].
[Arg1 The price of bananas] was increased again [Arg0 by Big Fruit Co.].
[Arg1 The price of bananas] increased [Arg2 5%].
Modifiers or adjuncts of the predicate: Arg-M

| ArgM-TMP  | when?                | yesterday evening, now |
| LOC       | where?               | at the museum, in San Francisco |
| DIR       | where to/from?       | down, to Bangkok |
| MNR       | how?                 | clearly, with much enthusiasm |
| PRP/CAU   | why?                 | because ... , in response to the ruling |
| REC       |                       | themselves, each other |
| ADV       | miscellaneous         | ...ate the meat raw |
| PRD       | secondary predication |                       |
Analysts have been expecting a GM-Jaguar pact that would give the U.S. car maker an eventual 30% stake in the British company.
The same parse tree PropBanked

expect(Analysts, GM-J pact)
give(GM-J pact, US car maker, 30% stake)
Annotated PropBank Data

- Penn English TreeBank, OntoNotes 5.0.
  - Total ~2 million words
- Penn Chinese TreeBank
- Hindi/Urdu PropBank
- Arabic PropBank

2013 Verb Frames Coverage
Count of word sense (lexical units)

<table>
<thead>
<tr>
<th>Language</th>
<th>Final Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>10,615*</td>
</tr>
<tr>
<td>Chinese</td>
<td>24,642</td>
</tr>
<tr>
<td>Arabic</td>
<td>7,015</td>
</tr>
</tbody>
</table>

From Martha Palmer 2013 Tutorial
Plus nouns and light verbs

Example Noun: Decision

- Roleset: Arg0: decider, Arg1: decision…

- “…[your\textsubscript{ARG0}] [decision\textsubscript{REL}]
  [to say look I don't want to go through this anymore\textsubscript{ARG1}]”

Example within an LVC: Make a decision

- “…[the President\textsubscript{ARG0}] [made\textsubscript{REL-LVB}]
  the [fundamentally correct\textsubscript{ARGM-ADJ}]
  [decision\textsubscript{REL}] [to get on offense\textsubscript{ARG1}]”

Slight from Palmer 2013
Semantic Role Labeling

FrameNet
Capturing descriptions of the same event by different nouns/verbs

\[
\text{[Arg}_1\text{ The price of bananas]} \text{ increased } \text{[Arg}_2\text{ 5\%].}
\]

\[
\text{[Arg}_1\text{ The price of bananas]} \text{ rose } \text{[Arg}_2\text{ 5\%].}
\]

There has been a \text{[Arg}_2\text{ 5\%] rise } \text{[Arg}_1\text{ in the price of bananas].}
FrameNet


• Roles in PropBank are specific to a verb

• Role in FrameNet are specific to a **frame**: a background knowledge structure that defines a set of frame-specific semantic roles, called **frame elements**,
  • includes a set of predicates that use these roles
  • each word evokes a frame and profiles some aspect of the frame
The “Change position on a scale” Frame

This frame consists of words that indicate the change of an ITEM’s position on a scale (the ATTRIBUTE) from a starting point (INITIAL VALUE) to an end point (FINAL VALUE).

- [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].
- [ITEM Colon cancer incidence] fell [DIFFERENCE by 50%] [GROUP among men].
- A steady increase [INITIAL_VALUE from 9.5] [FINAL_VALUE to 14.3] [ITEM in dividends]
- A [DIFFERENCE 5%] [ITEM dividend] increase...
### The “Change position on a scale” Frame

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

**ADVERBS:**
- decrease
diminish
dip
double
drop

**NOUNS:**
- hike
- increase
- rise
The “Change position on a scale” Frame

Core Roles

<table>
<thead>
<tr>
<th>ATTRIBUTE</th>
<th>The ATTRIBUTE is a scalar property that the ITEM possesses.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIFFERENCE</td>
<td>The distance by which an ITEM changes its position on the scale.</td>
</tr>
<tr>
<td>FINAL_STATE</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>FINAL_VALUE</td>
<td>The position on the scale where the ITEM ends up.</td>
</tr>
<tr>
<td>INITIAL_STATE</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>INITIAL_VALUE</td>
<td>The initial position on the scale from which the ITEM moves away.</td>
</tr>
<tr>
<td>ITEM</td>
<td>The entity that has a position on the scale.</td>
</tr>
<tr>
<td>VALUE_RANGE</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>DURATION</th>
<th>The length of time over which the change takes place.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPEED</td>
<td>The rate of change of the VALUE.</td>
</tr>
<tr>
<td>GROUP</td>
<td>The GROUP in which an ITEM changes the value of an ATTRIBUTE in a specified way.</td>
</tr>
</tbody>
</table>
Relation between frames

Inherits from:
Is Inherited by:
Perspective on:
Is Perspectivized in:
Uses:
Is Used by:
Subframe of:
Has Subframe(s):
Precedes:
Is Preceded by:
Is Inchoative of:
Is Causative of:
Relation between frames

“cause change position on a scale”

Is Causative of: **Change_position_on_a_scale**

Adds an agent Role

\[\text{AGENT They} \text{ raised} \text{ ITEM the price of their soda} \text{ DIFFERENCE by 2%}.\]

Relations between frames

Figure from Das et al 2010
1. Introduction

FrameNet (Fillmore, Johnson, and Petruck 2003) is a linguistic resource storing considerable information about lexical and predicate-argument semantics in English. Grounded in the theory of frame semantics (Fillmore 1982), it suggests—but does not formally define—a semantic representation that blends representations familiar from word-sense disambiguation (Ide and Véronis 1998) and semantic role labeling (SRL; Gildea and Jurafsky 2002). Given the limited size of available resources, accurately producing richly structured frame-semantic structures with high coverage will require data-driven techniques beyond simple supervised classification, such as latent variable modeling, semi-supervised learning, and joint inference.

In this article, we present a computational and statistical model for frame-semantic parsing, the problem of extracting from text semantic predicate-argument structures such as those shown in Figure 1. We aim to predict a frame-semantic representation with two statistical models rather than a collection of local classifiers, unlike earlier approaches (Baker, Ellsworth, and Erk 2007). We use a probabilistic framework that cleanly integrates the FrameNet lexicon and limited available training data. The probabilistic framework we adopt is highly amenable to future extension through new features, more relaxed independence assumptions, and additional semi-supervised models.

Carefully constructed lexical resources and annotated data sets from FrameNet, detailed in Section 3, form the basis of the frame structure prediction task. We decompose this task into three subproblems: target identification (Section 4), in which frame-evoking predicates are marked in the sentence; frame identification (Section 5), in which the evoked frame is selected for each predicate; and argument identification (Section 6), in which arguments to each frame are identified and labeled with a role from that frame. Experiments demonstrating favorable performance to the previous state of the art on SemEval 2007 and FrameNet data sets are described in each section. Some novel aspects of our approach include a latent-variable model (Section 5.2) and a semi-supervised extension of the predicate lexicon (Section 5.5) to facilitate disambiguation of words not in the FrameNet lexicon; a unified model for finding and labeling arguments.
FrameNet Complexity

But there still are n't enough ringers to ring more than six of the eight bells.

From Das et al. 2010
FrameNet and PropBank representations

In that time more than 1.2 million jobs have been created and the official jobless rate has been pushed below 17% from 21%.
Semantic Role Labeling

Algorithm
Semantic role labeling (SRL)

- The task of finding the semantic roles of each argument of each predicate in a sentence.
- FrameNet versus PropBank:

  [You] can’t [blame] [the program] [for being unable to identify it]
  COGNIZER TARGET EVALUEN REASON

  [The San Francisco Examiner] issued [a special edition] [yesterday]
  ARG0 TARGET ARG1 ARGM-TMP
History

• Semantic roles as a intermediate semantics, used early in
  • machine translation (Wilks, 1973)
  • question-answering (Hendrix et al., 1973)
  • spoken-language understanding (Nash-Webber, 1975)
  • dialogue systems (Bobrow et al., 1977)

• Early SRL systems

Simmons 1973, Marcus 1980:
  • parser followed by hand-written rules for each verb
  • dictionaries with verb-specific case frames (Levin 1977)
Why Semantic Role Labeling

• A useful shallow semantic representation

• Improves NLP tasks like:
  • question answering
    Shen and Lapata 2007, Surdeanu et al. 2011
  • machine translation
    Liu and Gildea 2010, Lo et al. 2013
Recall that the difference between these two models of semantic roles is that FrameNet employs many frame-specific frame elements as roles, while PropBank uses a smaller number of numbered argument labels that can be interpreted as verb-specific labels, along with the more general ARGM labels. Some examples:

\[(22.27)\]

\[
\begin{array}{ll}
\text{[You]} & \text{can't} \\
\text{[blame]} & \text{[the program]} \\
\text{[for being unable to identify it]} \\
\end{array}
\]

\[
\text{COGNIZER TARGET EV ALUEE REASON}
\]

\[(22.28)\]

\[
\begin{array}{ll}
\text{[The San Francisco Examiner]} & \text{issued} \\
\text{[a special edition]} & \text{[yesterday]} \\
\end{array}
\]

\[
\text{ARG} \quad \text{ARG} \quad \text{ARGM} \quad \text{-} \quad \text{TMP}
\]

A simplified semantic role labeling algorithm is sketched in Fig. 22.4. While there are a large number of algorithms, many of them use some version of the steps in this algorithm. Most algorithms, beginning with the very earliest semantic role analyzers (Simmons, 1973), begin by parsing, using broad-coverage parsers to assign a parse to the input string. Figure 22.5 shows a parse of (22.28) above. The parse is then traversed to find all words that are predicates. For each of these predicates, the algorithm examines each node in the parse tree and decides the semantic role (if any) it plays for this predicate. This is generally done by supervised classification. Given a labeled training set such as PropBank or FrameNet, a feature vector is extracted for each node, using feature templates described in the next subsection. A 1-of-N classifier is then trained to predict a semantic role for each constituent given these features, where N is the number of potential semantic roles plus an extra NONE role for non-role constituents. Most standard classification algorithms have been used (logistic regression, SVM, etc). Finally, for each test sentence to be labeled, the classifier is run on each relevant constituent. We give more details of the algorithm after we discuss features.

**function** SEMANTICROLELABEL \((\text{words})\) **returns** labeled tree

\[
\begin{align*}
parse & \leftarrow \text{PARSE}(\text{words}) \\
\text{for each} \hspace{0.2cm} \text{predicate} \hspace{0.2cm} \text{in} \hspace{0.2cm} \text{parse} \hspace{0.2cm} \text{do} \\
\hspace{1cm} \text{for each} \hspace{0.2cm} \text{node} \hspace{0.2cm} \text{in} \hspace{0.2cm} \text{parse} \hspace{0.2cm} \text{do} \\
\hspace{2cm} \text{featurevector} & \leftarrow \text{EXTRACTFEATURES}(\text{node}, \text{predicate}, \text{parse}) \\
\hspace{2cm} \text{CLASSIFYNODE}(\text{node}, \text{featurevector}, \text{parse})
\end{align*}
\]
How do we decide what is a predicate

- If we’re just doing PropBank verbs
  - Choose all verbs
  - Possibly removing light verbs (from a list)
- If we’re doing FrameNet (verbs, nouns, adjectives)
  - Choose every word that was labeled as a target in training data
Semantic Role Labeling

Figure 22.5 Parse tree for a PropBank sentence, showing the PropBank argument labels. The dotted line shows the path feature NP-SBJ = ARG0, the VP constituent.

- The headword of the constituent, Examiner.
- The headword part of speech, NNP.
- The path in the parse tree from the constituent to the predicate. This path is marked by the dotted line in Fig. 22.5. Following Gildea and Jurafsky (2000), we can use a simple linear representation of the path, NP-SBJ = ARG0.

- The voice of the clause in which the constituent appears, in this case, active (as contrasted with passive). Passive sentences tend to have strongly different linkings of semantic roles to surface form than do active ones.

- The binary linear position of the constituent with respect to the predicate, either before or after.

- The subcategorization of the predicate, the set of expected arguments that appear in the verb phrase. We can extract this information by using the phrase-structure rule that expands the immediate parent of the predicate; VP ! VBD NP PP for the predicate in Fig. 22.5.

- The named entity type of the constituent.

- The first words and the last word of the constituent.

The following feature vector thus represents the first NP in our example (recall that most observations will have the value NONE rather than, for example, ARG0, since most constituents in the parse tree will not bear a semantic role):

ARG0: [issued, NP, Examiner, NNP, NP-SBJ = ARG0, active, before, VP ! NP PP, ORG, The, Examiner]

Other features are often used in addition, such as sets of n-grams inside the constituent, or more complex versions of the path features (the upward or downward halves, or whether particular nodes occur in the path).

It's also possible to use dependency parses instead of constituency parses as the basis of features, for example using dependency parse paths instead of constituency paths.
Features

Headword of constituent
Examiner

Headword POS
NNP

Voice of the clause
Active

Subcategorization of pred
VP -> VBD NP PP

Named Entity type of constit
ORGANIZATION

First and last words of constit
The, Examiner

Linear position, clause re: predicate
before
Path Features

Path in the parse tree from the constituent to the predicate

\[ \text{NP} \uparrow \text{S} \downarrow \text{VP} \downarrow \text{VBD} \]
## Frequent path features

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Path</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.2%</td>
<td>VB↑VP↓PP</td>
<td>PP argument/adjunct</td>
</tr>
<tr>
<td>11.8</td>
<td>VB↑VP↑S↓NP</td>
<td>subject</td>
</tr>
<tr>
<td>10.1</td>
<td>VB↑VP↓NP</td>
<td>object</td>
</tr>
<tr>
<td>7.9</td>
<td>VB↑VP↑VP↑S↓NP</td>
<td>subject (embedded VP)</td>
</tr>
<tr>
<td>4.1</td>
<td>VB↑VP↓ADVP</td>
<td>adverbial adjunct</td>
</tr>
<tr>
<td>3.0</td>
<td>NN↑NP↑NP↓PP</td>
<td>prepositional complement of noun</td>
</tr>
<tr>
<td>1.7</td>
<td>VB↑VP↓PRT</td>
<td>adverbial particle</td>
</tr>
<tr>
<td>1.6</td>
<td>VB↑VP↑VP↑VP↑S↓NP</td>
<td>subject (embedded VP)</td>
</tr>
<tr>
<td>14.2</td>
<td></td>
<td>no matching parse constituent</td>
</tr>
<tr>
<td>31.4</td>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

For the purposes of choosing a frame element label for a constituent, the path feature is similar to the governing category feature defined above. Because the path captures more information, it may be more susceptible to parser errors and data sparseness. As an indication of this, the path feature is...
Final feature vector

• For “The San Francisco Examiner”,
• Arg0, [issued, NP, Examiner, NNP, active, before, VP → NP PP, ORG, The, Examiner, NP↑S↓VP↓VBD ]

• Other features could be used as well
  • sets of n-grams inside the constituent
  • other path features
    • the upward or downward halves
    • whether particular nodes occur in the path
3-step version of SRL algorithm

1. **Pruning**: use simple heuristics to prune unlikely constituents.
2. **Identification**: a binary classification of each node as an argument to be labeled or a NONE.
3. **Classification**: a 1-of-$N$ classification of all the constituents that were labeled as arguments by the previous stage.
Why add Pruning and Identification steps?

- Algorithm is looking at one predicate at a time
- Very few of the nodes in the tree could possibly be arguments of that one predicate
- Imbalance between
  - positive samples (constituents that are arguments of predicate)
  - negative samples (constituents that are not arguments of predicate)
- Imbalanced data can be hard for many classifiers
- So we prune the very unlikely constituents first, and then use a classifier to get rid of the rest.

- Add sisters of the predicate, then aunts, then great-aunts, etc
  - But ignoring anything in a coordination structure

- Machine learning for semantic role labeling

  - In addition, since it is not uncommon for a constituent to be assigned multiple semantic roles by different predicates (generally a predicate can only assign one semantic role to a constituent), the semantic role labeling system can only look at one predicate at a time, trying to find all the arguments for this particular predicate in the tree. The tree will be traversed as many times as there are predicates in the tree. This means there is an even higher proportion of constituents in the parse tree that are not arguments for the predicate the semantic role labeling system is currently looking at any given point. There is thus a serious imbalance between positive samples (constituents that are arguments to a particular predicate) and negative samples (constituents that are not arguments to this particular predicate). Machine learning algorithms generally do not handle extremely unbalanced data very well.

  - For these reasons, many systems divide the semantic role labeling task into two steps, identification, in which a binary decision is made as to whether a constituent carries a semantic role for a given predicate, and classification in which the specific semantic role is chosen. Separate machine learning classifiers are trained for these two tasks, often with many of the same features (Gildea and Jurafsky, 2002; Pradhan et al., 2005).

  - Another approach is to use a set of heuristics to prune out the majority of the negative samples, as a predicate's roles are generally found in a limited number of syntactic relations to the predicate itself. Some semantic labeling systems use a combination of both approaches: heuristics are first applied to prune out the constituents that are obviously not an argument for a certain predicate, and then a binary classifier is trained to further separate the positive samples from the negative samples. The goal of this filtering process is just to decide whether a constituent is an argument or not. Then a multi-class classifier is trained to decide the specific semantic role for this argument.

  - In the filtering stage, it is generally a good idea to be conservative and err on the side of keeping too many constituents rather than being too aggressive and filtering out true arguments. This can be achieved by lowering the threshold for positive samples, or conversely, raising the threshold for negative samples.
A common final stage: joint inference

• The algorithm so far classifies everything *locally* – each decision about a constituent is made independently of all others.

• But this can’t be right: Lots of global or joint interactions between arguments:
  • Constituents in FrameNet and PropBank must be non-overlapping.
    • A local system may incorrectly label two overlapping constituents as arguments.
  • PropBank does not allow multiple identical arguments
    • labeling one constituent ARG0
    • Thus should increase the probability of another being ARG1
How to do joint inference

• Reranking
  • The first stage SRL system produces multiple possible labels for each constituent
  • The second stage classifier the best global label for all constituents
  • Often a classifier that takes all the inputs along with other features (sequences of labels)
More complications: FrameNet

We need an extra step to find the frame

```plaintext
function SEMANTIC_ROLE_LABEL(words) returns labeled tree
  parse ← PARSE(words)
  for each predicate in parse do
    PredicateVector ← ExtractFrameFeatures(predicate, parse)
    Frame ← ClassifyFrame(predicate, PredicateVector)
  for each node in parse do
    for each feature in Frame do
      featureVector ← ExtractFeatures(node, predicate, parse, feature)
      Node ← ClassifyNode(node, featureVector, parse, Frame)
```

Dan Jurafsky
Features for Frame Identification

Das et al (2014)

the POS of the parent of the head word of \( t_i \)
the set of syntactic dependencies of the head word\(^{21} \) of \( t_i \)
if the head word of \( t_i \) is a verb, then the set of dependency labels of its children
the dependency label on the edge connecting the head of \( t_i \) and its parent
the sequence of words in the prototype, \( w_\ell \)
the lemmatized sequence of words in the prototype
the lemmatized sequence of words in the prototype and their part-of-speech tags \( \pi_\ell \)
WordNet relation\(^{22} \) \( \rho \) holds between \( \ell \) and \( t_i \)
WordNet relation\(^{22} \) \( \rho \) holds between \( \ell \) and \( t_i \), and the prototype is \( \ell \)
WordNet relation\(^{22} \) \( \rho \) holds between \( \ell \) and \( t_i \), the POS tag sequence of \( \ell \) is \( \pi_\ell \), and the POS tag sequence of \( t_i \) is \( \pi_t \)

---

\(^{21}\) If the target is not a subtree in the parse, we consider the words that have parents outside the span, and apply three heuristic rules to select the head: (1) choose the first word if it is a verb; (2) choose the last word if the first word is an adjective; (3) if the target contains the word \( \text{of} \), and the first word is a noun, we choose it. If none of these hold, choose the last word with an external parent to be the head.

\(^{22}\) These are: IDENTICAL-WORD, SYNONYM, ANTONYM (including extended and indirect antonyms), HYPERNYM, HYPONYM, DERIVED FORM, MORPHOLOGICAL VARIANT (e.g., plural form), VERB GROUP, ENTAILMENT, ENTAILED-BY, SEE-ALSO, CAUSAL RELATION, and NO RELATION.
4.3. LANGUAGE-(IN)DEPENDENT SEMANTIC ROLE LABELING

Commonalities

Like English semantic role labeling, Chinese semantic role labeling can be formulated as a classification task with three distinct stages: pruning, argument identification, and argument classification. The pruning algorithm described in Chapter 3 turns out to be straightforward to implement for Chinese data, and it involves minor changes in the phrase labels. For example, IP in the Chinese Treebank corresponds roughly to S in the Penn Treebank, and CP corresponds roughly to SBAR. Example 23 illustrates how the pruning algorithm works for Chinese. Assuming the predicate of interest is 调查 ("investigate"), the algorithm first adds the NP (事故 "accident" 原因 "cause") to the list of candidates. Then it moves up a level and adds the two ADVPs (正在 "now" 详细 "thoroughly") to the list of candidates. At the next level, the two VPs form a coordination structure and thus no candidate is added. Finally, at the next level, the NP (警方 "police") is added to the list of candidates. Obviously, the pruning algorithm works better when the parse trees that are the input to the semantic role labeling system are correct. In a realistic scenario, the parse trees are generated by a syntactic parser and are not expected to be perfect. However, experimental results

“...police are thoroughly investigating the cause of the accident.”
Not just verbs: NomBank

Meyers et al. 2004

Figure from Jiang and Ng 2006
Additional Issues for nouns

• Features:
  • Nominalization lexicon (employment $\rightarrow$ employ)
  • Morphological stem
    • Healthcare, Medicate $\rightarrow$ care

• Different positions
  • Most arguments of nominal predicates occur inside the NP
  • Others are introduced by support verbs
  • Especially light verbs “X made an argument”, “Y took a nap”
Semantic Role Labeling

Conclusion
Semantic Role Labeling

- A level of shallow semantics for representing events and their participants
  - Intermediate between parses and full semantics
- Two common architectures, for various languages
  - FrameNet: frame-specific roles
  - PropBank: Proto-roles
- Current systems extract by
  - parsing sentence
  - Finding predicates in the sentence
  - For each one, classify each parse tree constituent