We describe an approach to frame-semantic role labeling and evaluate it on data from this task.
SEMAFOR: Frame Argument Resolution with Log-Linear Models

or, The Case of the Missing Arguments

Desai Chen    Nathan Schneider    Dipanjan Das    Noah A. Smith

(guy in the front of the room)

We describe an approach to frame-semantic role labeling and evaluate it on data from this task.
Holmes sprang in his chair as if he had been stung when I read the headline.

This is a full annotation of a sentence in terms of its frames/arguments. Note that this is a *partial* semantic representation: it shows a certain amount of relational meaning but doesn’t encode, for instance, that “as if he had been stung” is a hypothetical used to provide imagery for the manner of motion (we infer that it must have been rapid and brought upon by a shocking stimulus).

The SRL task: Given a sentence with POS tags, syntactic dependencies, predicates, and frame names, predict the arguments for each frame role.

New wrinkle in this version of the task: classifying and resolving missing arguments.
Holmes *sprang* in his chair as if he had been *stung* when I *read* the headline.

This is a full annotation of a sentence in terms of its frames/arguments. Note that this is a *partial* semantic representation: it shows a certain amount of relational meaning but doesn’t encode, for instance, that “as if he had been stung” is a hypothetical used to provide imagery for the manner of motion (we infer that it must have been rapid and brought upon by a shocking stimulus).

The SRL task: Given a sentence with POS tags, syntactic dependencies, predicates, and frame names, predict the arguments for each frame role.

New wrinkle in this version of the task: classifying and resolving missing arguments.
This is a full annotation of a sentence in terms of its frames/arguments. Note that this is a *partial* semantic representation: it shows a certain amount of relational meaning but doesn’t encode, for instance, that “as if he had been stung” is a hypothetical used to provide imagery for the manner of motion (we infer that it must have been rapid and brought upon by a shocking stimulus).

The SRL task: Given a sentence with POS tags, syntactic dependencies, predicates, and frame names, predict the arguments for each frame role.

New wrinkle in this version of the task: classifying and resolving missing arguments.
Holmes *sprang* in his chair as if he had been *stung* when I *read* the headline.

This is a full annotation of a sentence in terms of its frames/arguments. Note that this is a *partial* semantic representation: it shows a certain amount of relational meaning but doesn’t encode, for instance, that “as if he had been stung” is a hypothetical used to provide imagery for the manner of motion (we infer that it must have been rapid and brought upon by a shocking stimulus).

The SRL task: Given a sentence with POS tags, syntactic dependencies, predicates, and frame names, predict the arguments for each frame role.

New wrinkle in this version of the task: classifying and resolving missing arguments.
Contributions

• Evaluate frame SRL on new data
• Experiment with a classifier for null instantiations (NIs)
  › implicit interactions in a discourse
Overview

Background: frame SRL

- Overt argument identification
- Null instantiation resolution
- Conclusion
FrameNet

- **FrameNet** (Fillmore et al., 2003) defines semantic frames, roles, and associated predicates
  - provides a linguistically **rich** representation for predicate-argument structures based on the theory of **frame semantics** (Fillmore, 1982)
The FrameNet lexicon is a repository of expert information, storing the semantic frames and a number of (frame-specific) roles. Each frame represents a holistic event or scenario, generalizing over specific predicates. It also defines roles for the participants, props, and attributes of the scenario.
For example, here we show the Make_noise frame that has several roles such as Sound, Noisy_event, Sound_Source, etc. FrameNet also lists some possible lexical units which could evoke these frames. Examples for this frame are cough, gobble, hiss, ring, and so on.
The FrameNet lexicon also provides relationships between frames and between roles

http://framenet.icsi.berkeley.edu
[SE’07] has ANC travel guides, PropBank news, and (mostly) NTI reports on weapons stockpiles. Unlike other participants, we do not use the 139,000 lexicographic exemplar sentences (except indirectly through features) because the annotations are partial (only 1 frame) and the sample of sentences is biased (they were chosen manually to illustrate variation of arguments).

[SE’10] also has coreference, though we do not make use of this information.
Annotated Data

- Full-text annotations: all frames + arguments
  - [SE’07] SemEval 2007 task data:
    - news, popular nonfiction, bureaucratic

[SE’07] has ANC travel guides, PropBank news, and (mostly) NTI reports on weapons stockpiles. Unlike other participants, we do not use the 139,000 lexicographic exemplar sentences (except indirectly through features) because the annotations are partial (only 1 frame) and the sample of sentences is biased (they were chosen manually to illustrate variation of arguments).

[SE’10] also has coreference, though we do not make use of this information.
Annotated Data

- Full-text annotations: all frames + arguments
  - **[SE’07]** SemEval 2007 task data:
    - news, popular nonfiction, bureaucratic
  - **[SE’10]** New SemEval 2010 data:
    - fiction

[SE’07] has ANC travel guides, PropBank news, and (mostly) NTI reports on weapons stockpiles. Unlike other participants, we do not use the 139,000 lexicographic exemplar sentences (except indirectly through features) because the annotations are partial (only 1 frame) and the sample of sentences is biased (they were chosen manually to illustrate variation of arguments).

[SE’10] also has coreference, though we do not make use of this information.
Overview

✓ Background: frame SRL

➡ Overt argument identification

• Null instantiation resolution

• Conclusion
Frame SRL: Overt Arguments

We train a **classifier** to pick an argument for each role of each frame.

(Das et al., 2010)

See NAACL 2010 paper
Frame SRL: Overt Arguments

We train a **classifier** to pick an argument for each role of each frame.

\[ \text{sprang} \quad \text{SELF\textunderscore MOTION} \_\text{Place} \]

\[ \text{in his chair} \]

\[ \text{in} \quad \text{PRP} \quad \text{NN} \]

\[ \emptyset \]

*a probabilistic model with features looking at the span, the frame, the role, and the observed sentence structure*

(Das et al., 2010)

See NAACL 2010 paper
An example of the desired mapping. For the predicate ‘sprang’, each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.
An example of the desired mapping. For the predicate ‘sprang’, each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.
A example of the desired mapping. For the predicate ‘sprang’, each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.
An example of the desired mapping. For the predicate ‘sprang’, each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.
An example of the desired mapping. For the predicate ‘sprang’, each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.
An example of the desired mapping. For the predicate ‘sprang’, each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.
...and likewise for ‘stung’, etc.
Frame SRL: Overt Arguments

*stung ~ EXPERIENCER_OBJ*

Experiencer

Stimulus Ø

Degree Ø

Time Ø

Manner Ø

... and likewise for ‘stung’, etc.
Frame SRL: Experimental Setup

• SRL component of SEMAFOR 1.0 (Das et al., 2010; http://www.ark.cs.cmu.edu/SEMAFOR)

• Task scoring script for overt argument precision, recall, $F_1$ on test set
  ‣ Strict matching criterion: argument spans must be exact
SE’07: SEMAFOR trained only on old data (different domain from test set)
SE’10: new training data included (same domain as test set)
Adding a small amount of new data helps a lot: (~7% F1): domain issue + so little data to begin with. Suggests even more data might yield substantial improvements!

Scores are microaveraged according to the number of frames in each of the 2 test documents.
Overview

✓ Background: frame SRL
✓ Overt argument identification
→ Null instantiation resolution
• Conclusion
Null Instantiations

• New this year: classification and resolution of null instantiations (NIs), arguments that are nonlocal or implicit in the discourse
  ‣ a role is said to be null-instantiated if it has no (overt) argument in the sentence, but has an implicit contextual filler
  ‣ see also (Gerber & Chai, 2010), which considers implicit argument resolution for several (nominal) predicates

(Fillmore, 1986; Ruppenhofer, 2005)
Null Instantiations

• indefinite null instantiation (INI): the referent is vague/deemphasized
  ‣ *We ate $\emptyset$Thing\_eaten*.
  ‣ *He was stung $\emptyset$Stimulus*.

(Fillmore, 1986; Ruppenhofer, 2005)
Null Instantiations

• **indefinite null instantiation (INI):** the referent is vague/deemphasized
  
  ‣ *We* ate $\emptyset$Thing\_eaten.
  
  ‣ *He* was *stung* $\emptyset$Stimulus.

• **definite null instantiation (DNI):** a *specific* referent is obvious from the discourse
  
  ‣ *They’ll* *arrive* soon $\emptyset$Goal.
  (the goal is implicitly the speaker’s location)

(Fillmore, 1986; Ruppenhofer, 2005)
“I think I shall be in a position to make the situation rather more clear to you before long. It has been an exceedingly difficult and most complicated business.

(SemEval 2010 test data)

The other frame-evoking words are bolded, but their arguments are not shown.
DNI Example: overt nonlocal referent

“I think I shall be in a position to make the situation rather more clear to you before long. It has been an exceedingly difficult and most complicated business ØExperiencer.

(SemEval 2010 test data)

The other frame-evoking words are bolded, but their arguments are not shown.
DNI Example: overt nonlocal referent

“I think I shall be in a position to make the situation rather more clear to you before long. It has been an exceedingly difficult and most complicated business.

(SemEval 2010 test data)

The other frame-evoking words are bolded, but their arguments are not shown.
These numbers may be approximate. They show how few NIs there are compared to overt args, and why the DNI resolution task is so hard.
We try a straightforward approach for null instantiations: a second classifier.

The SRL module selects an argument span or none for each role. For core roles, we then build a second classifier for disambiguating types of null elements. This uses the same mathematical techniques to predict a different kind of outputs.

Ideally, the NI module would be able to predict whether each core role was INI, DNI + its referent, if applicable, or not NI. Our system only considers DNIs with referents in the previous 3 sentences. Experiments show that a large search space, while leading to high *oracle* recall, confuses the model in practice.
We try a straightforward approach for null instantiations: a second classifier.

The SRL module selects an argument span or none for each role. For core roles, we then build a second classifier for disambiguating types of null elements. This uses the same mathematical techniques to predict a different kind of outputs.

Ideally, the NI module would be able to predict whether each core role was INI, DNI + its referent, if applicable, or not NI. Our system only considers DNIs with referents in the previous 3 sentences. Experiments show that a large search space, while leading to high *oracle* recall, confuses the model in practice.
We try a straightforward approach for null instantiations: a second classifier.

The SRL module selects an argument span or none for each role. For core roles, we then build a second classifier for disambiguating types of null elements. This uses the same mathematical techniques to predict a different kind of outputs. Ideally, the NI module would be able to predict whether each core role was INI, DNI + its referent, if applicable, or not NI. Our system only considers DNI s with referents in the previous 3 sentences. Experiments show that a large search space, while leading to high *oracle* recall, confuses the model in practice.
Evaluating NI performance only. We train only on the new SemEval 2010 data because the SemEval 2007 data used different annotation practices for null instantiations.

The evaluation criterion actually doesn’t distinguish between INIs and unresolved DNIs. We predicted only the former.
Overview

✓ Background: frame SRL
✓ Overt argument identification
✓ Null instantiation resolution
➡ Conclusion
Contributions & Claims

1. Evaluated frame SRL on **new data**
   - Amount of training data makes a big difference
   - Still lots of room for improvement

2. Experimented with a classifier for **null instantiations**, with mixed success
   - Resolving nonlocal referents is much harder than classifying the instantiation type

3. Learned models achieve higher **recall**, and consequently $F_1$, than custom heuristics used by other teams
   - Our modeling framework is **extensible**: it should allow us to incorporate many of these in a soft way as features
Sizes of frame-annotated data provided for SemEval ’07 and ’10 tasks, as compared to PropBank. The bottom graph is in terms of tokens. Whereas FrameNet provides a linguistically rich representation, PropBank has much higher coverage/annotated data.
Sizes of frame-annotated data provided for SemEval ’07 and ’10 tasks, as compared to PropBank. The bottom graph is in terms of tokens. Whereas FrameNet provides a linguistically rich representation, PropBank has much higher coverage/annotated data.
Sizes of frame–annotated data provided for SemEval ’07 and ’10 tasks, as compared to PropBank. The bottom graph is in terms of tokens. Whereas FrameNet provides a linguistically rich representation, PropBank has much higher coverage/annotated data.
Sizes of frame-annotated data provided for SemEval ’07 and ’10 tasks, as compared to PropBank. The bottom graph is in terms of tokens. Whereas FrameNet provides a linguistically rich representation, PropBank has much higher coverage/annotated data.
Conclusion

• Next challenge: **data sparseness** in frame SRL
  ‣ obtaining quality frame annotations from experts is expensive
  ‣ opportunity for semi-supervised learning
  ‣ additional knowledge/constraints in modeling
  ‣ non-expert annotations?
  ‣ bridging across lexical-semantic resources (FrameNet, WordNet, PropBank, VerbNet, NomBank, …)
Task 10 (Frame SRL) Posters

(101) CLR: Linking Events and Their Participants in Discourse Using a Comprehensive FrameNet Dictionary
Ken Litkowski

(102) VENSES++: Adapting a deep semantic processing system to the identification of null instantiations
Sara Tonelli & Rodolfo Delmonte

if you’re interested in this task…
Thank you!

Image from http://commons.wikimedia.org/wiki/File:SherlockHolmes.jpg
Thank you!
• NI subtask confusion matrix
• NI-only and full results table


### Table 3.

**NI-only Subtask: Confusion Matrix**

<table>
<thead>
<tr>
<th>Gold</th>
<th>overt</th>
<th>DNI</th>
<th>INI</th>
<th>masked</th>
<th>inc.</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>overt</td>
<td>2068</td>
<td>5</td>
<td>362</td>
<td>327</td>
<td>0</td>
<td>2762</td>
</tr>
<tr>
<td>DNI</td>
<td>64</td>
<td>12</td>
<td>182</td>
<td>90</td>
<td>0</td>
<td>348</td>
</tr>
<tr>
<td>INI</td>
<td>41</td>
<td>2</td>
<td>214</td>
<td>96</td>
<td>0</td>
<td>353</td>
</tr>
<tr>
<td>masked</td>
<td>73</td>
<td>0</td>
<td>240</td>
<td>1394</td>
<td>0</td>
<td>1707</td>
</tr>
<tr>
<td>inc.</td>
<td>12</td>
<td>2</td>
<td>55</td>
<td>2</td>
<td>0</td>
<td>71</td>
</tr>
<tr>
<td>total</td>
<td>2258</td>
<td>21</td>
<td>1053</td>
<td>1909</td>
<td>0</td>
<td>3688 correct</td>
</tr>
</tbody>
</table>

From the paper
Instances with explicit referents will not even be considered in the training data but at a cost. This narrows the search space considerably for DNI referents. Consider nouns, pronouns, and noun phrases from outside the sentence along with a special category. These non-local candidate fillers are handled differently from candidates within the sentence. How are they selected using more restrictive criteria? This suggests that data sparseness is hindering our system's ability to learn useful generalizations about NIs. The worse the trained model performs at distinguishing NIs from non-NIs, the more candidate DNI referents are under consideration. Experiments with variants on these assumptions show that the larger the search space, the more candidate DNI referents are under consideration.

Investigation of separate modeling is left to future work. In practice, we found that these features receive negligible weight and had virtually no effect on performance, possibly due to data sparseness. An additional change in the feature set is that ordering-distance features (Das et al.) encode the maximum distributional similarity to any word heading a filler of that role in the example.Modifiers thus limit oracle recall to about 75% of DNIs. The first indicates whether the head word is from the target sentence. The second indicates whether the head word is the thing eaten in the sentence. The thing eaten in the sentence is handled as a non-local phrase. For each candidate span, we use two types of features. The modified feature set includes path features with simpler features derived from FrameNet's lexicographic exemplar annotations. These features thus limit oracle recall to about 75% of DNIs.

### Results Table: NI-only and Full

<table>
<thead>
<tr>
<th>NI-only</th>
<th>Training Data</th>
<th>Chapter 13</th>
<th>Chapter 14</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec.</td>
<td>Rec.</td>
<td>$F_1$</td>
</tr>
<tr>
<td>SemEval 2010 new: 100%</td>
<td>0.40</td>
<td>0.64</td>
<td>0.50</td>
</tr>
<tr>
<td>SemEval 2010 new: 75%</td>
<td>0.66</td>
<td>0.37</td>
<td>0.50</td>
</tr>
<tr>
<td>SemEval 2010 new: 50%</td>
<td>0.73</td>
<td>0.38</td>
<td>0.51</td>
</tr>
<tr>
<td>Full</td>
<td>0.35</td>
<td>0.55</td>
<td>0.43</td>
</tr>
</tbody>
</table>

**Chapter 13**

- **SemEval 2010 new: 100%**
  - Prec.: 0.40
  - Rec.: 0.64
  - $F_1$: 0.50

**Chapter 14**

- **SemEval 2010 new: 100%**
  - Prec.: 0.53
  - Rec.: 0.60
  - $F_1$: 0.56