Frame Semantic Structure Extraction

Organizing team: Collin Baker, Michael Ellsworth
(International Computer Science Institute, Berkeley),
Katrin Erk(U Texas, Austin)

October 4, 2006

1 Description of task

This task will consist of automatically recognizing expressions that evoke semantic frames, discriminating the word sense (frame) of each evoking expression, and labeling its syntactic dependents (broadly construed) with regard to which roles in that frame they fill. The frames and roles and the training data will be provided by the FrameNet project. Some syntactic information will be supplied in the training and testing data, including POS tagging, phrase type of the role fillers, and at least one parse of the texts. Participants will be free to use any sort of NLP to achieve the results, but the output of the participating systems will be a simplified dependency graph representing only the semantic relations. Thus, this task is related to (but more demanding than) the Automatic Semantic Role Labeling task of Senseval-3, and closer to applications such as text summarization and information extraction.

1.1 FrameNet

The Berkeley FrameNet project is creating a computer- and human-readable lexical resource for English, based on the theory of frame semantics and supported by corpus evidence. The aim is to document the range of semantic and syntactic combinatory possibilities (valences) of each word in each of its senses, through computer-assisted annotation of example sentences and automatic tabulation and display of the annotation results. The current release of the FrameNet data, which is freely available for instructional and research purposes, includes roughly 780 frames with roughly 10,000 word senses (lexical units), more than 6,000 of which have been fully exemplified in annotated sentences.

The ultimate goal is to represent the lexical semantics of all the sentences in a text, based on the relations between predicators and their dependents, including both phrases and clauses, which may, in turn, include other predicators. For more information and to view the FrameNet data, please visit `http://framenet.icsi.berkeley.edu`

1.2 Semantic parsers

Since the seminal publication of Gildea and Jurafsky (2002), many researchers have been interested in building software that will automatically recognize the senses of predicators and label their dependents with semantic roles; such systems have been referred to as semantic parsers or automatic semantic role labelers, and have used a great variety of training features and learning algorithms, as exemplified in the semantic role labeling tasks in Senseval-3 and CoNLL-2004 and 2005.

We are aware of at least two semantic role labelers that have been trained on the FrameNet data. One, called Shalmaneser, was developed by Katrin Erk and Sebastian Padó as part of the SALSA project at the University of the Saarland, and can be trained on either the FrameNet data for English or the SALSA data for German. The other, ASSERT, developed by Sameer Pradhan and others at University of Colorado has been used primarily with data from the PropBank project, but can also be trained on FrameNet data. At least one other semantic parser using FrameNet categories is being developed by Ana-Maria Giuglea and Alessandro Moschitti and they hope to release it publicly soon.

2 Training data

The training data for the task will consist of the new data release from FrameNet (Release 1.3). This contains roughly 150 K annotation sets, of which 139 K are lexicographic examples, with each sentence annotated for a single predicator. The remainder are from full-text annotation in which each sentence is annotated for all predicators; 1,700 sentences are annotated in the full-text portion of the database, accounting for roughly 11,700 annotation sets, or 6.8 predicators (=annotation sets) per sentence. Both the lexicographic and the full-text data can be used to train recognizers which are then run to produce full-text annotation.

3 Testing data

The texts for the testing data will be taken from the American National Corpus (ANC) Version 2, and will include roughly 300 sentences. These will be fully annotated, and will contain roughly 2,000 annotation sets.

4 Evaluation criteria

Two types of evaluation are proposed:

4.1 Label matching evaluation:

This type of evaluation would be similar to that used in the SensEval-3 Semantic labeling task, except that stricter standards would apply regarding boundary identification.

4.2 Semantic dependency evaluation:

Since the role fillers are dependents (broadly speaking) of the predicators, the full FrameNet annotation of a sentence is roughly equivalent to a dependency parse, in which some of the arcs are labeled with role names; and a dependency graph can be derived algorithmically from FrameNet annotation.

For example, consider the sentence (from the ANC), "Have you ever seen an old photo of yourself and been embarrassed at the way you looked?". The frames and frame elements are as follows:

Frame	**FE (role)**	**evoker/filler**
Perception_experience	Evoked by	*see.v*
	Perceiver_passive	*you*
	Phenomenon	*an old photo of yourself*
Emotion_directed	evoked by	*embarrassed.a*
	Experiencer	*you*
	Stimulus	*at the way you look*
Appearance	evoked by	*look.v*
	Phenomenon	*you*
	Appraisal	*the way*
Physical_artworks	evoked by	*photo.n*
	Descriptor	*old*
	Artifact	*photo*
	Represented	*of yourself*

The corresponding XML used for the evaluation will look like this:

```
<W ID=W1 start=0 end=3 form="Have"  />
<W ID=W2 start=5 end=7 form="you"  />
<W ID=W3 start=9 end=12 form="ever"  />
<W ID=W4 start=14 end=17 form="seen"  />
<W ID=W5 start=19 end=20 form="an"  />
<W ID=W6 start=22 end=24 form="old"  />
<W ID=W7 start=26 end=30 form="photo"  />
<W ID=W8 start=32 end=33 form="of"  />
<W ID=W9 start=35 end=42 form="yourself"  />
```

```
<W ID=W10 start=44 end=46 form="and"  />
<W ID=W11 start=48 end=51 form="been"  />
<W ID=W12 start=53 end=63 form="embarrassed"  />
<W ID=W13 start=65 end=66 form="at"  />
<W ID=W14 start=68 end=70 form="the"  />
<W ID=W15 start=72 end=74 form="way"  />
<W ID=W16 start=76 end=78 form="you"  />
<W ID=W17 start=80 end=86 form="looked"  />
<W ID=W18 start=81 end=81 form="?"  />

<P ID=P1  >
  <E Head=T Ref=P2 />
  <E Ref=W10 /> <!-- and -->
  <E Head=T Ref=P3 />
</P>

<P ID=P2 Frame="Perception_experience" >
  <E FE="Perceiver_passive" Ref=W2 />
  <E Target=T Ref=W4 /> <!-- seen -->
  <E FE="Phenomenon" Ref=P4 />
</p>

<P ID=P3 >
  <E Cop=T Ref=W11 /> <!-- been -->
  <E Head=T Ref=P7 />
</P>

<P ID=P4 >
  <E Ref=W5 /> <!-- an -->
  <E Head=T Ref=P5 />
</P>

<P ID=P5 Frame="Physical_artworks" >
  <E FE="Descriptor"  Ref=W6 /> <!-- old -->
  <E Target=T Ref=W7 /> <!-- photo -->
  <E FE="Artifact" Ref=W7 />
  <E FE="Represented" Ref=P6 />
</P>

<P ID=P6 >
  <E Ref=W8 /> <!-- of -->
  <E Head=T Ref=W9 /> <!-- yourself -->
</P>

<P ID=P7 Frame="Emotion_directed" >
  <E FE="Experiencer" Ref=W2 />
  <E Target=T Ref=W12 /> <!-- embarrassed -->
  <E FE="Stimulus" Ref=P8 />
</P>

<P ID=P8 >
  <E Ref=W13 /> <!-- at -->
  <E Head=T Ref=P10 />
</P>

<P ID=P9 >
  <E Ref=W14 /> <!-- the -->
  <E Head=T Ref=W15 /> <!-- way -->
```

```
</P>

<P ID=P10 Frame="Appearance" >
  <E FE="Appraisal" Ref=P9 />
  <E FE="Phenomenon" Ref=W16 /> <!-- you -->
  <E Target=T Ref=W17 /> <!-- looked -->
</P>
```

A graphical view equivalent to the above XML is given at the end of this paper. The semantic dependency evaluation will score the participants' output on the basis of how well it agrees with the gold standard with regard to this dependency tree structure. For this purpose, for non-frame nodes, a node will count as matching provided that it includes the head of the gold standard, whether or not non-head children of that node are included. For nodes which represent evocations of frames, the participants will get full credit if the frame of the node matches the gold standard. If the frame in the answer is different from but related to the gold standard frame via frame-to-frame relations, partial credit will be given, decreasing exponentially according to the distance (number of links) in the frame-frame relation graph. Each frame element link must match the gold standard frame element and contain at least the same head word in order to gain full credit; again, partial credit will be given for frame elements related via FE-to-FE relations.

For both types of evaluation, credit will be given for reasonable "extensions" to the training data, by judicious guessing as to words that were not covered in the training data (i.e. Release 1.3). As the FrameNet staff annotates the testing data, in cases where appropriate frames and word senses (LUs) do not exist, they will be created in the course of the annotation process, as is usual. Participants who are able to approximate these new data by guessing correctly as to what existing frame the new word sense would be in or by placing it in a closely related frame in the frame hierarchy will get credit for doing so. ("Closely related" can be measured in terms of links in the frame hierarchy.)

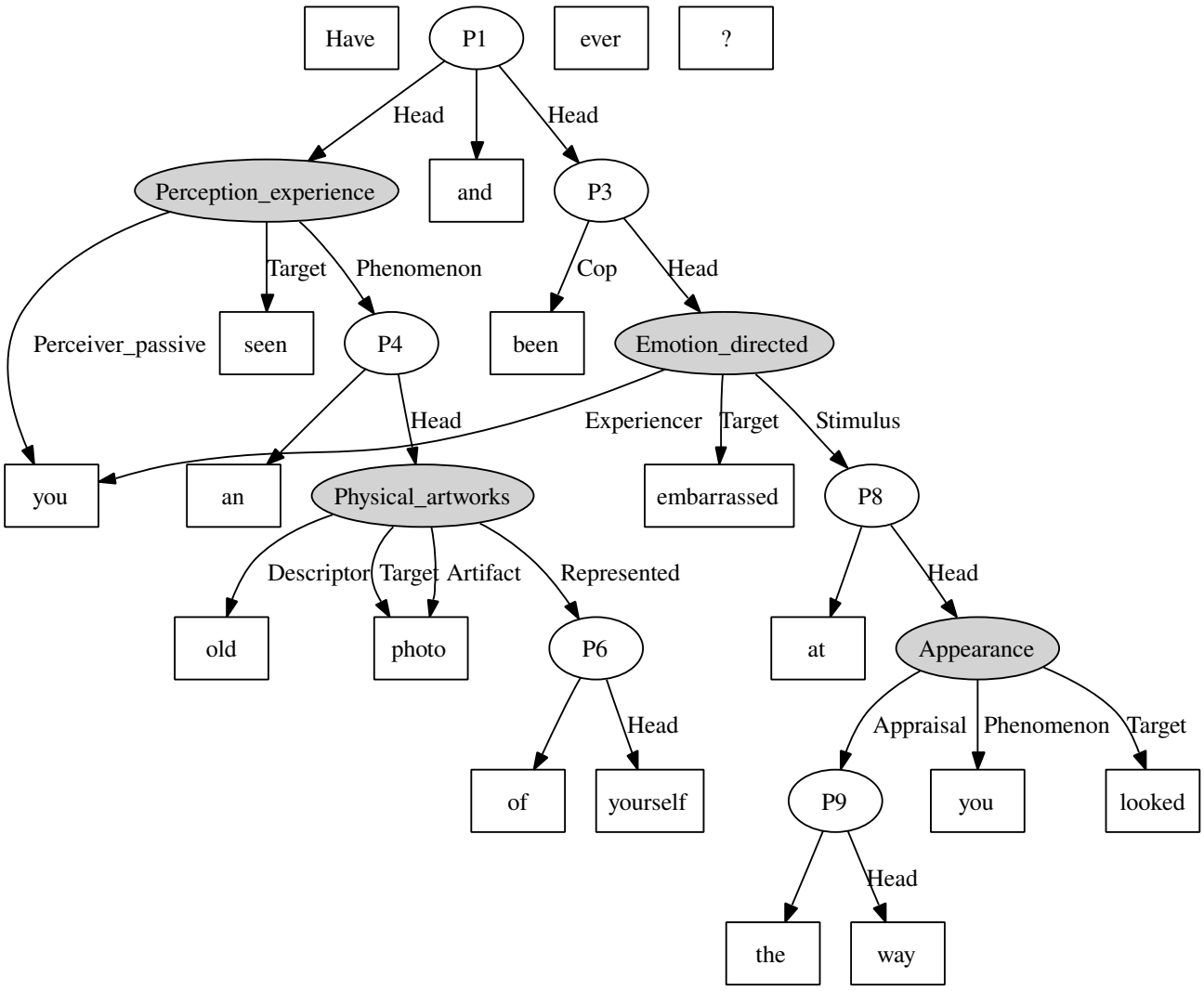
5 Availability of resources to participants

All the data is available in XML with DTDs; tools are also provided to convert from XML into OWL DL. A script will also be provided to convert standard FN full-text XML into the simpler semantic dependency representation that will be used for evaluation.

- Release 1.3 of the FrameNet data is available for research purposes at no cost from the FrameNet project (`http://framenet.icsi.berkeley.edu`).
- All of the annotated data for this task will be from the "redistributable" portion of the ANC, and so can be provided free of charge to the participants. The full Release 2 of the ANC is available from the Linguistic Data Consortium (LDC) (`http://www.ldc.upenn.edu`); the cost is $75 for non-members.
- Shalmaneser is available at no cost from the SALSA project (`http://www.coli.uni-saarland.de/projects/salsa/shal/`).
- ASSERT is available at no cost from U Colorado (`http://oak.colorado.edu/assert/`).)

6 Resources needed to prepare task

We believe the FrameNet team will be able to prepare this data in the normal course of our work on current projects.


Have
P1
ever
?
Head
Head
Perception_experience
and
P3
Target
Phenomenon
Cop
Head
Perceiver_passive
seen
P4
been
Emotion_directed
Head
Experiencer
Target
Stimulus
you
an
Physical_artworks
embarrassed
P8
Descriptor
Target
Artifact
Represented
Head
old
photo
P6
at
Appearance
Head
Appraisal
Phenomenon
Target
of
yourself
P9
you
looked
Head
the
way