CoNLL Shared Task 2010: Sentiment Proposal

1 Organizing committee

- Dan Jurafsky, Stanford University.
- Lillian Lee, Cornell University.
- Jim Martin, University of Colorado.
- Nicolas Nicolov, J.D.Power and Associates, McGraw-Hill. [contact person]
- Martha Palmer, University of Colorado.
- Janyce Wiebe, University of Pittsburgh.

2 Task

The proposed staged tasks are to determine the author’s sentiment towards an entity based on:

- **Stage 1**: A few (possibly named) mentions of the same entity (which could be seen as matches of a search query). Ex: “Audi R8” occurrences in a text document.
- **Stage 2**: Stage 1 + coreference mention on that top entity. This group of mentions constitutes the top-level entity. Ex: “it”, “the supercar”;
- **Stage 3**: Stage 2 + meronymy mentions and PART-OF links between meronymy mentions that ultimately lead to a mention of the top-level entity;
- **Stage 4**: Stage 3 + all sentiment expressions (without orientation), intensifiers/shifters, negators, etc. in the entire document but not links between an expression and its target.

The documents are user-generated content (blogs), and the semantic types are products (currently vehicles and cameras; we are adding other domains — cell phones, camcorders).

To do this sentiment task well, it is useful for systems to be able to perform named entity recognition (which in turn benefits from chunking), dependency parsing and semantic role labeling. These support tasks have been addressed in previous CoNLL evaluations and, thus, the CoNLL community is a natural group of experts to address the new sentiment task.

To allow teams to participate that do not have a large suite of NLP components we propose to consider the stage tasks.

No restrictions are placed on the kind of systems participating teams can develop and there is no requirement that the additional information provided in the different stages be used (even though we do believe it will help).

3 Dataset

**Description of annotation.** This corpus has been annotated for entity mentions, meronymy, coreference and fine-grained expression of sentiment. Entity mentions are references in the text to entities such as persons, organizations (e.g., companies), times, locations, units (e.g., price, age), vehicles (i.e., trucks, SUVs, cars) and vehicle parts and features. Meronymy relations involve linking parts of an entity to the whole (e.g., “The car . . . the engine . . .”). Coreference annotation is the linking of text spans that refer to the same entity (e.g., “I love my [Subaru Forrester]. [It] drives well.”). Sentiment expressions are text spans that indicate positive or negative sentiment being expressed about a particular entity mention (e.g., “This car is [great].” “I [love] this car.”). Furthermore, the following categories are identified and labeled to capture the precise source and intensity of the sentiment: committer (e.g., “definitely”); intensifier (e.g., “very”); comparison (e.g., “X gets better Y than Z”); OtherPersonsOpinion (e.g., “John thinks . . .”); Neutralizer (e.g., “if”); Negator (e.g., “not”, “never”).

Each entity has also been annotated for overall sentiment using the following categories:

- **POSITIVE** (“great camera!”);
- **NEGATIVE** (“limited battery life”);
- **MIXED** (“I love the way the car handles but the engine isn’t that powerful.”);
- **NEUTRAL** — the author is expressing an opinion on the entity but it is neither positive nor negative (“The stereo is ok.”);
- **UNKNOWN** — there is no sentiment expressed associated with any of the mentions of the entity.

Additionally the data (including the test set) is tokenized, part-of-speech tagged (using SVMTag), and parsed (using the MaltParser).
**Data acquisition.** The raw texts were acquired by carrying out web searches. Subsequently a human judgement on the suitability for inclusion in the corpus was performed. Each document has associated metadata: id, uri, datetime, authorUri, gender, age, location, vertical, sourceType (blog or message board), blogType (personal, professional, or community), guestPost (true or false).

**Annotation tool.** The annotation tool used for this project is Knowtator ([http://knowtator.sourceforge.net](http://knowtator.sourceforge.net)), a Java-based plug-in built on the Protégé platform. Knowtator distinguishes between classes, which are categories that can be assigned to text spans (e.g., EntityMention; SentimentBearingExpression; Negator; Neutralizer; Committer), and slots in those classes, which express links between class members (e.g., the PartOf slot; the RefersTo slot, which expresses coreference between two entity mentions; or the Target slot, which expresses the target of the sentiment of a SentimentBearingExpression). Knowtator uses standoff offset annotation which is represented in XML format.

**Corpus statistics.** The corpus currently consists of 462 annotated texts in the automotive and camera domains. 7 of those texts have been triple annotated, and 34 double annotated.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>462</td>
</tr>
<tr>
<td>Tokens</td>
<td>263,277</td>
</tr>
<tr>
<td>Sentences</td>
<td>16,455</td>
</tr>
<tr>
<td>Mentions</td>
<td>56,194</td>
</tr>
<tr>
<td>Entities</td>
<td>37,365</td>
</tr>
<tr>
<td>Meronyms</td>
<td>105,969</td>
</tr>
<tr>
<td>Entities with coref./meronyms</td>
<td>4,416</td>
</tr>
<tr>
<td>ITA car mentions</td>
<td>93.14%</td>
</tr>
<tr>
<td>ITA car-part mentions</td>
<td>92.46%</td>
</tr>
<tr>
<td>ITA target relation</td>
<td>91.04%</td>
</tr>
<tr>
<td>ITA sent. expr. extent</td>
<td>81.76%</td>
</tr>
<tr>
<td>ITA sent. expr. polarity</td>
<td>90.50%</td>
</tr>
</tbody>
</table>

We are adding the calculation of the interannotator agreement at the entity level to the evaluation script. At the current speed of annotation we will provide 300K (or more) token training set by end of December 2009 (our internal target is at 330k). We will also provide software APIs for reading and writing the data in Knowtator XML format.

**Test set.** The test set will contain documents from the training domains as well as new domains (e.g., TVs, game consoles) to test portability of systems. We are aiming to provide 50K tokens of test data.

**Availability.** The data is freely available for research purposes. Participants will be required to sign a non-commercial license.

### 4 Evaluation

**Input:** A list of triples:  
1. Gold annotation file,
2. One mention from the initially provided ones in Stage 1 above,
3. System output file

**Output:** Accuracy of identifying entity sentiment.

The evaluation is the same for each of the stages. It is expected the systems will score better when provided with more information. We have evaluation software such as Python scripts which we will make available.

**Additional software.** Software APIs for reading and writing the data in Knowtator XML format, visualization of annotations in GraphViz format.

**Multilinguality.** We have received substantial interest from colleagues working on other languages and can coordinate the creation of a corresponding multilingual dataset for a CoNLL shared task in subsequent years.

**Acknowledgements.** The J.D.Power and Associates team working on this corpus consists of: Dr. Miriam Eckert, Steliana Ivanova, Jason Kessler (also at Indiana University), Ron Woodward, and five annotators. We would like to thank Prof. Mike Mozer (Univ. of Colorado) for useful discussions as part of the design and annotation of the sentiment corpus.

Proposals should be sent to conll10st@stp.lingfil.uu.se no later than Mon, 2009-09-21.