Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns

Yejin Choi and Claire Cardie
Department of Computer Science
Cornell University
Ithaca, NY 14853
{ychoi,cardie}@cs.cornell.edu

Ellen Riloff and Siddharth Patwardhan
School of Computing
University of Utah
Salt Lake City, UT 84112
{riloff,sidd}@cs.utah.edu

Abstract

Recent systems have been developed for sentiment classification, opinion recognition, and opinion analysis (e.g., detecting polarity and strength). We pursue another aspect of opinion analysis: identifying the sources of opinions, emotions, and sentiments. We view this problem as an information extraction task and adopt a hybrid approach that combines Conditional Random Fields (Lafferty et al., 2001) and a variation of AutoSlog (Riloff, 1996a). While CRFs model source identification as a sequence tagging task, AutoSlog learns extraction patterns. Our results show that the combination of these two methods performs better than either one alone. The resulting system identifies opinion sources with 79.3% precision and 59.5% recall using a head noun matching measure, and 81.2% precision and 60.6% recall using an overlap measure.

1 Introduction

In recent years, there has been a great deal of interest in methods for automatically identifying opinions, emotions, and sentiments in text. Much of this research explores sentiment classification, a text categorization task in which the goal is to classify a document as having positive or negative polarity (e.g., Das and Chen (2001), Pang et al. (2002), Turney (2002), Dave et al. (2003), Pang and Lee (2004)). Other research efforts analyze opinion expressions at the sentence level or below to recognize opinions, their polarity, and their strength (e.g., Dave et al. (2003), Pang and Lee (2004), Wilson et al. (2004), Yu and Hatzivassiloglou (2003), Wiebe and Riloff (2005)). Many applications could benefit from these opinion analyzers, including product reputation tracking (e.g., Morinaga et al. (2002), Yi et al. (2003)), opinion-oriented summarization (e.g., Cardie et al. (2004)), and question answering (e.g., Bethard et al. (2004), Yu and Hatzivassiloglou (2003)).

We focus here on another aspect of opinion analysis: automatically identifying the sources of the opinions. Identifying opinion sources will be especially critical for opinion-oriented question-answering systems (e.g., systems that answer questions of the form "How does [X] feel about [Y]?") and opinion-oriented summarization systems, both of which need to distinguish the opinions of one source from those of another.[1]

The goal of our research is to identify direct and indirect sources of opinions, emotions, sentiments, and other *private states* that are expressed in text. To illustrate the nature of this problem, consider the examples below:

S1: <u>Taiwan-born voters</u> favoring independence...

---

[1] In related work, we investigate methods to identify the opinion expressions (e.g., Riloff and Wiebe (2003), Wiebe and Riloff (2005), Wilson et al. (2005)) and the nesting structure of sources (e.g., Breck and Cardie (2004)). The *target* of each opinion, i.e., what the opinion is directed towards, is currently being annotated manually for our corpus.

S2: According to the report, the human rights record in China is horrendous.

S3: International officers believe that the EU will prevail.

S4: International officers said US officials want the EU to prevail.

In S1, the phrase "Taiwan-born voters" is the direct (i.e., first-hand) source of the "favoring" sentiment. In S2, "the report" is the direct source of the opinion about China's human rights record. In S3, "International officers" are the direct source of an opinion regarding the EU. The same phrase in S4, however, denotes an indirect (i.e., second-hand, third-hand, etc.) source of an opinion whose direct source is "US officials".

In this paper, we view *source identification* as an information extraction task and tackle the problem using sequence tagging and pattern matching techniques simultaneously. Using syntactic, semantic, and orthographic lexical features, dependency parse features, and opinion recognition features, we train a linear-chain Conditional Random Field (CRF) (Lafferty et al., 2001) to identify opinion sources. In addition, we employ features based on automatically learned extraction patterns and perform feature induction on the CRF model.

We evaluate our hybrid approach using the NRRC corpus (Wiebe et al., 2005), which is manually annotated with direct and indirect opinion source information. Experimental results show that the CRF model performs well, and that both the extraction patterns and feature induction produce performance gains. The resulting system identifies opinion sources with 79.3% precision and 59.5% recall using a head noun matching measure, and 81.2% precision and 60.6% recall using an overlap measure.

2 The Big Picture

The goal of information extraction (IE) systems is to extract information about events, including the participants of the events. This task goes beyond Named Entity recognition (e.g., Bikel et al. (1997)) because it requires the recognition of role relationships. For example, an IE system that extracts information about corporate acquisitions must distinguish between the company that is doing the acquiring and the company that is being acquired. Sim-

ilarly, an IE system that extracts information about terrorism must distinguish between the person who is the perpetrator and the person who is the victim. We hypothesized that IE techniques would be well-suited for source identification because an opinion statement can be viewed as a kind of speech event with the source as the agent.

We investigate two very different learning-based methods from information extraction for the problem of opinion source identification: graphical models and extraction pattern learning. In particular, we consider Conditional Random Fields (Lafferty et al., 2001) and a variation of AutoSlog (Riloff, 1996a). CRFs have been used successfully for Named Entity recognition (e.g., McCallum and Li (2003), Sarawagi and Cohen (2004)), and AutoSlog has performed well on information extraction tasks in several domains (Riloff, 1996a). While CRFs treat source identification as a sequence tagging task, AutoSlog views the problem as a pattern-matching task, acquiring symbolic patterns that rely on both the syntax and lexical semantics of a sentence. We hypothesized that a combination of the two techniques would perform better than either one alone.

Section 3 describes the CRF approach to identifying opinion sources and the features that the system uses. Section 4 then presents a new variation of AutoSlog, *AutoSlog-SE*, which generates IE patterns to extract sources. Section 5 describes the hybrid system: we encode the IE patterns as additional features in the CRF model. Finally, Section 6 presents our experimental results and error analysis.

3 Semantic Tagging via Conditional Random Fields

We defined the problem of opinion source identification as a sequence tagging task via CRFs as follows. Given a sequence of tokens, $x = x_1x_2...x_n$, we need to generate a sequence of tags, or labels, $y = y_1y_2...y_n$. We define the set of possible label values as $'S'$, $'T'$, $'-'$, where $'S'$ is the first token (or Start) of a source, $'T'$ is a non-initial token (i.e., a conTinuation) of a source, and $'-'$ is a token that is not part of any source.[2]

A detailed description of CRFs can be found in

---

[2] This is equivalent to the IOB tagging scheme used in syntactic chunkers (Ramshaw and Marcus, 1995).

Lafferty et al. (2001). For our sequence tagging problem, we create a linear-chain CRF based on an undirected graph $G = (V, E)$, where V is the set of random variables $Y = \{Y_i | 1 \leq i \leq n\}$, one for each of n tokens in an input sentence; and $E = \{(Y_{i-1}, Y_i) | 1 < i \leq n\}$ is the set of $n - 1$ edges forming a linear chain. For each sentence x, we define a non-negative clique potential $\exp(\sum_{k=1}^{K} \lambda_k f_k(y_{i-1}, y_i, x))$ for each edge, and $\exp(\sum_{k=1}^{K'} \lambda'_k f'_k(y_i, x))$ for each node, where $f_k(...)$ is a binary feature indicator function, λ_k is a weight assigned for each feature function, and K and K' are the number of features defined for edges and nodes respectively. Following Lafferty et al. (2001), the conditional probability of a sequence of labels y given a sequence of tokens x is:

$$P(y|x) = \frac{1}{Z_x} \exp\left(\sum_{i,k} \lambda_k f_k(y_{i-1}, y_i, x) + \sum_{i,k} \lambda'_k f'_k(y_i, x)\right) \tag{1}$$

$$Z_x = \sum_y \exp\left(\sum_{i,k} \lambda_k f_k(y_{i-1}, y_i, x) + \sum_{i,k} \lambda'_k f'_k(y_i, x)\right) \tag{2}$$

where Z_x is a normalization constant for each x. Given the training data D, a set of sentences paired with their correct `'ST-'` source label sequences, the parameters of the model are trained to maximize the conditional log-likelihood $\prod_{(x,y) \in D} P(y|x)$. For inference, given a sentence x in the test data, the tagging sequence y is given by $\text{argmax}_{y'} P(y'|x)$.

3.1 Features

To develop features, we considered three properties of opinion sources. First, the sources of opinions are mostly noun phrases. Second, the source phrases should be semantic entities that can bear or express opinions. Third, the source phrases should be directly related to an opinion expression. When considering only the first and second criteria, this task reduces to named entity recognition. Because of the third condition, however, the task requires the recognition of opinion expressions and a more sophisticated encoding of sentence structure to capture relationships between source phrases and opinion expressions.

With these properties in mind, we define the following features for each token/word x_i in an input sentence. For pedagogical reasons, we will describe some of the features as being multi-valued or categorical features. In practice, however, all features are binarized for the CRF model.

Capitalization features We use two boolean features to represent the capitalization of a word: `all-capital`, `initial-capital`.

Part-of-speech features Based on the lexical categories produced by GATE (Cunningham et al., 2002), each token x_i is classified into one of a set of coarse part-of-speech tags: noun, verb, adverb, wh-word, determiner, punctuation, etc. We do the same for neighboring words in a $[-2, +2]$ window in order to assist noun phrase segmentation.

Opinion lexicon features For each token x_i, we include a binary feature that indicates whether or not the word is in our *opinion lexicon* — a set of words that indicate the presence of an opinion. We do the same for neighboring words in a $[-1, +1]$ window. Additionally, we include for x_i a feature that indicates the opinion subclass associated with x_i, if available from the lexicon. (e.g., "bless" is classified as "moderately subjective" according to the lexicon, while "accuse" and "berate" are classified more specifically as "judgments".) The lexicon is initially populated with approximately 500 opinion words [3] from (Wiebe et al., 2002), and then augmented with opinion words identified in the training data. The training data contains manually produced phrase-level annotations for all expressions of opinions, emotions, etc. (Wiebe et al., 2005). We collected all content words that occurred in the training set such that at least 50% of their occurrences were in opinion annotations.

Dependency tree features For each token x_i, we create features based on the parse tree produced by the Collins (1999) dependency parser. The purpose of the features is to (1) encode structural information, and (2) indicate whether x_i is involved in any grammatical relations with an opinion word. Two pre-processing steps are required before features can be constructed:

---

[3] Some words are drawn from Levin (1993); others are from Framenet lemmas (Baker et al. 1998) associated with communication verbs.

1. **Syntactic chunking.** We traverse the dependency tree using breadth-first search to identify and group syntactically related nodes, producing a flatter, more concise tree. Each syntactic "chunk" is also assigned a grammatical role (e.g., `subject`, `object`, `verb modifier`, `time`, `location`, `of-pp`, `by-pp`) based on its constituents. Possessives (e.g., "Clinton's idea") and the phrase "according to X" are handled as special cases in the chunking process.

2. **Opinion word propagation.** Although the opinion lexicon contains only content words and no multi-word phrases, actual opinions often comprise an entire phrase, e.g., *"is really willing"* or *"in my opinion"*. As a result, we mark as an opinion the entire chunk that contains an opinion word. This allows each token in the chunk to act as an opinion word for feature encoding.

After syntactic chunking and opinion word propagation, we create the following dependency tree features for each token x_i:

- the grammatical role of its chunk
- the grammatical role of x_{i-1}'s chunk
- whether the parent chunk includes an opinion word
- whether x_i's chunk is in an argument position with respect to the parent chunk
- whether x_i represents a constituent boundary

Semantic class features We use 7 binary features to encode the semantic class of each word x_i: `authority`, `government`, `human`, `media`, `organization_or_company`, `proper_name`, and `other`. The `other` class captures 13 semantic classes that cannot be sources, such as `vehicle` and `time`.

Semantic class information is derived from named entity and semantic class labels assigned to x_i by the Sundance shallow parser (Riloff, 2004). Sundance uses named entity recognition rules to label noun phrases as belonging to named entity classes, and assigns semantic tags to individual words based on a semantic dictionary. Table 1 shows the hierarchy that Sundance uses for semantic classes associated with opinion sources. Sundance is also used to recognize and instantiate the source extraction patterns
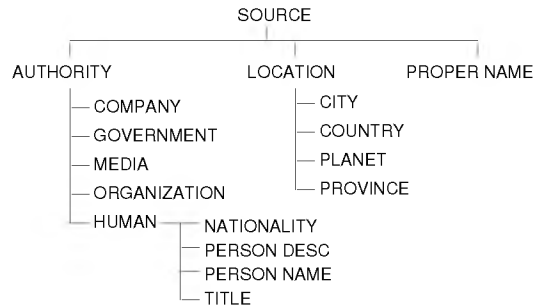

Figure 1: The semantic hierarchy for opinion sources

that are learned by AutoSlog-SE, which is described in the next section.

4 Semantic Tagging via Extraction Patterns

We also learn patterns to extract opinion sources using a statistical adaptation of the AutoSlog IE learning algorithm. AutoSlog (Riloff, 1996a) is a supervised extraction pattern learner that takes a training corpus of texts and their associated answer keys as input. A set of heuristics looks at the context surrounding each answer and proposes a lexico-syntactic pattern to extract that answer from the text. The heuristics are not perfect, however, so the resulting set of patterns needs to be manually reviewed by a person.

In order to build a fully automatic system that does not depend on manual review, we combined AutoSlog's heuristics with statistics from the annotated training data to create a fully automatic supervised learner. We will refer to this learner as AutoSlog-SE (Statistically Enhanced variation of AutoSlog). AutoSlog-SE's learning process has three steps:

Step 1: AutoSlog's heuristics are applied to every noun phrase (NP) in the training corpus. This generates a set of extraction patterns that, collectively, can extract every NP in the training corpus.

Step 2: The learned patterns are augmented with selectional restrictions that semantically constrain the types of noun phrases that are legitimate extractions for opinion sources. We used

the semantic classes shown in Figure 1 as selectional restrictions.

Step 3: The patterns are applied to the training corpus and statistics are gathered about their extractions. We count the number of extractions that match annotations in the corpus (correct extractions) and the number of extractions that do not match annotations (incorrect extractions). These counts are then used to estimate the probability that the pattern will extract an opinion source in new texts:

$$P(\text{source} \mid \text{pattern}_i) = \frac{\text{correct sources}}{\text{correct sources} + \text{incorrect sources}}$$

This learning process generates a set of extraction patterns coupled with probabilities. In the next section, we explain how these extraction patterns are represented as features in the CRF model.

5 Extraction Pattern Features for the CRF

The extraction patterns provide two kinds of information. `SourcePatt` indicates whether a word activates any source extraction pattern. For example, the word *"complained"* activates the pattern *"<subj> complained"* because it anchors the expression. `SourceExtr` indicates whether a word is extracted by any source pattern. For example, in the sentence "President Jacques Chirac frequently complained about France's economy", the words "President", "Jacques", and "Chirac" would all be extracted by the *"<subj> complained"* pattern.

Each extraction pattern has frequency and probability values produced by AutoSlog-SE, hence we create four IE pattern-based features for each token x_i: `SourcePatt-Freq`, `SourceExtr-Freq`, `SourcePatt-Prob`, and `SourceExtr-Prob`, where the frequency values are divided into three ranges: $\{0, 1, 2+\}$ and the probability values are divided into five ranges of equal size.

6 Experiments

We used the Multi-Perspective Question Answering (MPQA) corpus[4] for our experiments. This corpus

consists of 535 documents that have been manually annotated with opinion-related information including direct and indirect sources. We used 135 documents as a tuning set for model development and feature engineering, and used the remaining 400 documents for evaluation, performing 10-fold cross validation. These texts are English language versions of articles that come from many countries and cover many topics.[5]

We evaluate performance using 3 measures: overlap match (OL), head match (HM), and exact match (EM). OL is a lenient measure that considers an extraction to be correct if it overlaps with any of the annotated words. HM is a more conservative measure that considers an extraction to be correct if its head matches the head of the annotated source. We report these somewhat loose measures because the annotators vary in where they place the exact boundaries of a source. EM is the strictest measure that requires an exact match between the extracted words and the annotated words. We use three evaluation metrics: recall, precision, and F-measure with recall and precision equally weighted.

6.1 Baselines

We developed three baseline systems to assess the difficulty of our task. *Baseline-1* labels as sources all phrases that belong to the semantic categories `authority`, `government`, `human`, `media`, `organization_or_company`, `proper_name`. Table 1 shows that the precision is poor, suggesting that the third condition described in Section 3.1 (opinion recognition) does play an important role in source identification. The recall is much higher but still limited due to sources that fall outside of the semantic categories or are not recognized as belonging to these categories. *Baseline-2* labels a noun phrase as a source if any of the following are true: (1) the NP is the subject of a verb phrase containing an opinion word, (2) the NP follows *"according to"*, (3) the NP contains a possessive and is preceded by an opinion word, or (4) the NP follows *"by"* and attaches to an opinion word. *Baseline-2*'s heuristics are designed to address the first and the third conditions in Section 3.1. Table 1 shows that *Baseline-2* is substantially better than *Baseline-1*. *Baseline-3*

[4]The MPQA corpus can be freely obtained at http://nrrc.mitre.org/NRRC/publications.htm.

[5]This data was obtained from the Foreign Broadcast Information Service (FBIS), a U.S. government agency.

		Recall	Prec	F1
Baseline-1	OL	77.3	28.8	42.0
	HM	71.4	28.6	40.8
	EM	65.4	20.9	31.7
Baseline-2	OL	62.4	60.5	61.4
	HM	59.7	58.2	58.9
	EM	50.8	48.9	49.8
Baseline-3	OL	49.9	72.6	59.2
	HM	47.4	72.5	57.3
	EM	44.3	58.2	50.3
Extraction Patterns	OL	48.5	81.3	60.8
	HM	46.9	78.5	58.7
	EM	41.9	70.2	52.5
CRF: basic features	OL	56.1	81.0	66.3
	HM	55.1	79.2	65.0
	EM	50.0	72.4	59.2
CRF: basic + IE pattern features	OL	59.1	82.4	68.9
	HM	58.1	80.5	67.5
	EM	52.5	73.3	61.2
CRF-FI: basic features	OL	57.7	80.7	67.3
	HM	56.8	78.8	66.0
	EM	51.7	72.4	60.3
CRF-FI: basic + IE pattern features	OL	60.6	81.2	69.4
	HM	59.5	79.3	68.0
	EM	54.1	72.7	62.0

Table 1: Source identification performance table

labels a noun phrase as a source if it satisfies both *Baseline-1* and *Baseline-2*'s conditions (this should satisfy all three conditions described in Section 3.1). As shown in Table 1, the precision of this approach is the best of the three baselines, but the recall is the lowest.

6.2 Extraction Pattern Experiment

We evaluated the performance of the learned extraction patterns on the source identification task. The learned patterns were applied to the test data and the extracted sources were scored against the manual annotations.[6] Table 1 shows that the extraction patterns produced lower recall than the baselines, but with considerably higher precision. These results show that the extraction patterns alone can identify

---

[6]These results were obtained using the patterns that had a probability > .50 and frequency > 1.

nearly half of the opinion sources with good accuracy.

6.3 CRF Experiments

We developed our CRF model using the MALLET code from McCallum (2002). For training, we used a Gaussian prior of 0.25, selected based on the tuning data. We evaluate the CRF using the *basic features* from Section 3, both with and without the IE pattern features from Section 5. Table 1 shows that the CRF with basic features outperforms all of the baselines as well as the extraction patterns, achieving an F-measure of 66.3 using the OL measure, 65.0 using the HM measure, and 59.2 using the EM measure. Adding the IE pattern features further increases performance, boosting recall by about 3 points for all of the measures and slightly increasing precision as well.

CRF with feature induction. One limitation of log-linear function models like CRFs is that they cannot form a decision boundary from conjunctions of existing features, unless conjunctions are explicitly given as part of the feature vector. For the task of identifying opinion sources, we observed that the model could benefit from conjunctive features. For instance, instead of using two separate features, HUMAN and PARENT-CHUNK-INCLUDES-OPINION-EXPRESSION, the conjunction of the two is more informative.

For this reason, we applied the CRF feature induction approach introduced by McCallum (2003). As shown in Table 1, where CRF-FI stands for the CRF model with feature induction, we see consistent improvements by automatically generating conjunctive features. The final system, which combines the basic features, the IE pattern features, and feature induction achieves an F-measure of 69.4 (recall=60.6%, precision=81.2%) for the OL measure, an F-measure of 68.0 (recall=59.5%, precision=79.3%) for the HM measure, and an F-measure of 62.0 (recall=54.1%, precision=72.7%) for the EM measure.

6.4 Error Analysis

An analysis of the errors indicated some common mistakes:

- Some errors resulted from error propagation in

our subsystems. Errors from the sentence boundary detector in GATE (Cunningham et al., 2002) were especially problematic because they caused the Collins parser to fail, resulting in no dependency tree information.

- Some errors were due to complex and unusual sentence structure, which our rather simple feature encoding for CRF could not capture well.

- Some errors were due to the limited coverage of the opinion lexicon. We failed to recognize some cases when idiomatic or vague expressions were used to express opinions.

Below are some examples of errors that we found interesting. Doubly underlined phrases indicate incorrectly extracted sources (either false positives or false negatives). Opinion words are singly underlined.

False positives:

(1) Actually, these three countries do have one common denominator, i.e., that their values and policies do not agree with those of the United States and none of them are on good terms with the United States.

(2) Perhaps this is why Fidel Castro has not spoken out against what might go on in Guantanamo.

In (1), "their values and policies" seems like a reasonable phrase to extract, but the annotation does not mark this as a source, perhaps because it is somewhat abstract. In (2), "spoken out" is negated, which means that the verb phrase does not bear an opinion, but our system failed to recognize the negation.

False negatives:

(3) And for this reason, too, they have a moral duty to speak out, as Swedish Foreign Minister Anna Lindh, among others, did yesterday.

(4) In particular, Iran and Iraq are at loggerheads with each other to this day.

Example (3) involves a complex sentence structure that our system could not deal with. (4) involves an uncommon opinion expression that our system did not recognize.

7 Related Work

To our knowledge, our research is the first to automatically identify opinion sources using the MPQA opinion annotation scheme. The most closely related work on opinion analysis is Bethard et al. (2004), who use machine learning techniques to identify propositional opinions and their holders (sources). However, their work is more limited in scope than ours in several ways. Their work only addresses propositional opinions, which are "localized in the propositional argument" of certain verbs such as "believe" or "realize". In contrast, our work aims to find sources for all opinions, emotions, and sentiments, including those that are not related to a verb at all. Furthermore, Berthard et al.'s task definition only requires the identification of direct sources, while our task requires the identification of both direct and indirect sources. Bethard et al. evaluate their system on manually annotated FrameNet (Baker et al., 1998) and Prop-Bank (Palmer et al., 2005) sentences and achieve 48% recall with 57% precision.

Our IE pattern learner can be viewed as a cross between AutoSlog (Riloff, 1996a) and AutoSlog-TS (Riloff, 1996b). AutoSlog is a supervised learner that requires annotated training data but does not compute statistics. AutoSlog-TS is a weakly supervised learner that does not require annotated data but generates coarse statistics that measure each pattern's correlation with relevant and irrelevant documents. Consequently, the patterns learned by both AutoSlog and AutoSlog-TS need to be manually reviewed by a person to achieve good accuracy. In contrast, our IE learner, AutoSlog-SE, computes statistics directly from the annotated training data, creating a fully automatic variation of AutoSlog.

8 Conclusion

We have described a hybrid approach to the problem of extracting sources of opinions in text. We cast this problem as an information extraction task, using both CRFs and extraction patterns. Our research is the first to identify both direct and indirect sources for all types of opinions, emotions, and sentiments.

Directions for future work include trying to increase recall by identifying relationships between opinions and sources that cross sentence boundaries, and relationships between multiple opinion expressions by the same source. For example, the fact that a coreferring noun phrase was marked as a source in one sentence could be a useful clue for extracting the source from another sentence. The probability or the strength of an opinion expression may also play a useful role in encouraging or suppressing source extraction.

9 Acknowledgments

References

C. Baker, C. Fillmore & J. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the COLING-ACL*.

S. Bethard, H. Yu, A. Thornton, V. Hativassiloglou & D. Jurafsky. 2004. Automatic extraction of opinion propositions and their holders. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text*.

D. Bikel, S. Miller, R. Schwartz & R. Weischedel. 1997. Nymble: A High-Performance Learning Name-Finder. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*.

E. Breck & C. Cardie. 2004. Playing the Telephone Game: Determining the Hierarchical Structure of Perspective and Speech Expressions. In *Proceedings of 20th International Conference on Computational Linguistics*.

C. Cardie, J. Wiebe, T. Wilson & D. Litman. 2004. Low-level annotations and summary representations of opinions for multiperspective QA. In *New Directions in Question Answering*. AAAI Press/MIT Press.

M. Collins. 1999. Head-driven Statistical Models for Natural Language Parsing. Ph.D. thesis, University of Pennsylvania.

H. Cunningham, D. Maynard, K. Bontcheva & V. Tablan. 2002. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*.

S. Das & M. Chen. 2001. Yahoo for amazon: Extracting market sentiment from stock message boards. In *Proceedings of the 8th Asia Pacific Finance Association Annual Conference*.

K. Dave, S. Lawrence & D. Pennock. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *International World Wide Web Conference*.

J. Lafferty, A. K. McCallum & F. Pereira. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of 18th International Conference on Machine Learning*.

B. Levin. 1993. English Verb Classes and Alternations: A Preliminary Investigation. University of Chicago Press.

A. K. McCallum. 2002. *MALLET: A Machine Learning for Language Toolkit*. http://mallet.cs.umass.edu.

A. K. McCallum. 2003. Efficiently Inducing Features of Conditional Random Fields. In *Conference on Uncertainty in Artificial Intelligence*.

A. K. McCallum & W. Li. 2003. Early Results for Named Entity Recognition with Conditional Random Fields, Feature Induction and Web-Enhanced Lexicons. In *Conference on Natural Language Learning*.

S. Morinaga, K. Yamanishi, K. Tateishi & T. Fukushima 2002. Mining Product Reputations on the Web. In *Proceedings of the 8th Internatinal Conference on Knowledge Discover and Data Mining*.

M. Palmer, D. Gildea & P. Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. In *Computational Linguistics* 31.

B. Pang, L. Lee & S. Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*.

B. Pang & L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*.

L. A. Ramshaw & M. P. Marcus. 1995. Nymble: A High-Performance Learning Name-Finder. In *Proceedings of the 3rd Workshop on Very Large Corpora*.

E. Riloff. 1996a. An Empirical Study of Automated Dictionary Construction for Information Extraction in Three Domains. In *Artificial Intelligence*, Vol. 85.

E. Riloff. 1996b. Automatically Generating Extraction Patterns from Untagged Text. In *Proceedings of the 13th National Conference on Artificial Intelligence*.

E. Riloff & J. Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceesings of 2003 Conference on Empirical Methods in Natural Language Processing*.

E. Riloff & W. Phillips. 2004. An Introduction to the Sundance and AutoSlog Systems Technical Report UUCS-04-015, School of Computing, University of Utah.

S. Sarawagi & W. W. Cohen. 2004. Semi-Markov Conditional Random Fields for Information Extraction *18th Annual Conference on Neural Information Processing Systems*.

P. Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*.

T. Wilson, J. Wiebe & R. Hwa. 2004. Just how mad are you? finding strong and weak opinion clauses. In *Proceedings of the 9th National Conference on Artificial Intelligence*.

T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff & S. Patwardhan. 2005. OpinionFinder: A system for subjectivity analysis. Demonstration Description in *Conference on Empirical Methods in Natural Language Processing*.

J. Yi, T. Nasukawa, R. Bunescu & W. Niblack. 2003. Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques. In *Proceedings of the 3rd IEEE International Conference on Data Mining*.

H. Yu & V. Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

J. Wiebe, E. Breck, C. Buckley, C. Cardie, P. Davis, B. Fraser, D. Litman, D. Pierce, E. Riloff & T. Wilson. 2002. NRRC Summer Workshop on Multiple-Perspective Question Answering: Final Report.

J. Wiebe & E. Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. *Sixth International Conference on Intelligent Text Processing and Computational Linguistics*.

J. Wiebe, T. Wilson & C. Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 1(2).