Identifying Collocations for Recognizing Opinions

Janyce Wiebe and **Theresa Wilson** and **Matthew Bell**
University of Pittsburgh
wiebe,twilson,mbell@cs.pitt.edu

Proc. ACL/EACL '01 Workshop on Collocation, Toulouse, France, July 2001

Abstract

Subjectivity in natural language refers to aspects of language used to express opinions and evaluations (Banfield, 1982; Wiebe, 1994). There are numerous applications for which knowledge of subjectivity is relevant, including genre detection, information extraction, and information retrieval. This paper shows promising results for a straightforward method of identifying collocational clues of subjectivity, as well as evidence of the usefulness of these clues for recognizing opinionated documents.

1 Introduction

Subjectivity in natural language refers to aspects of language used to express opinions and evaluations (Banfield, 1982; Wiebe, 1994). *Subjectivity tagging* is distinguishing sentences used to present opinions and other forms of subjectivity (*subjective sentences*) from sentences used to objectively present factual information (*objective sentences*). This task is especially relevant for news reporting and Internet forums, in which opinions of various agents are expressed.

There are numerous applications for which subjectivity tagging is relevant. Two are information retrieval and information extraction. Current extraction and retrieval technology focuses almost exclusively on the subject matter of documents. However, additional aspects of a document influence its relevance, including, e.g., the evidential status of the material presented, and the attitudes expressed about the topic (Kessler et al., 1997). Knowledge of subjective language would also be useful in flame recognition (Spertus, 1997; Kaufer, 2000), email classification (Aone et al., 2000), intellectual attribution in text (Teufel and Moens, 2000), recognizing speaker role in radio broadcasts (Barzilay et al., 2000), review mining (Terveen et al., 1997), generation and style (Hovy, 1987), clustering documents by ideological point of view (Sack, 1995), and any other application that would benefit from knowledge of how opinionated the language is, and whether or not the writer purports to objectively present factual material.

To use subjectivity tagging in applications, good linguistic clues must be found. As with many pragmatic and discourse distinctions, existing lexical resources such as machine readable dictionaries (Procter, 1978) and ontologies for natural language processing (Hovy, 1998), while useful, are not sufficient for identifying such linguistic clues, because they are not comprehensively coded for subjectivity. The focus of this work is learning subjectivity clues from corpora. The goal is to develop a repository of linguistic clues of subjectivity and knowledge about the use of such clues in context, to enable NLP researchers to exploit subjective language in applications such as those mentioned above.

This paper takes a step in that direction by demonstrating a straightforward method for automatically identifying collocational clues of subjectivity in texts. The method is first used to identify collocations composed of fixed sequences of words (stems) which, when they appear together, tend to be subjective. These include expressions such as 'village idiot' and 'get out of here'. (Smajda, 1993) argues for the importance of identifying more general forms of collocation, and this paper addresses one unusual type of generalization: a position in the collocation can be filled by any word that appears infrequently in the corpus. There are more than expected low-frequency words in subjective sentences; apparently people are creative when they are being opinionated. Rather than discarding these low-frequency words, this paper provides evidence that using them in the context of collocations can be informative for recognizing subjectivity. Both the fixed and generalized types of collocation show increased precision, and their performance is consistent with the performance of previously identified clues of subjectivity (Wiebe, 2000). In addition, preliminary results suggest that they are useful for identifying opinionated documents such as editorials.

In the remainder of this paper, subjectivity and previous work on automatic recognition of subjectivity are first described. The training data is then described, followed by the method for extracting collocations. The experiments and results are presented next, followed by conclusions and future work.

2 Subjectivity

Sentence (1) is an example of a simple subjective sentence, and (2) is an example of a simple objective sentence:[1]

(1) At several different layers, it's a fascinating tale.

(2) Bell Industries Inc. increased its quarterly to 10 cents from 7 cents a share.

The main types of subjectivity are:

1. *Evaluation.* This category includes emotions such as hope and hatred as well as evaluations, judgements, and opinions. Examples of expressions involving positive evaluation are 'enthused', 'wonderful', and 'great product!'. Examples involving negative evaluation are 'complained', 'you idiot!', and 'terrible product'.

2. *Speculation*. This category includes anything that removes the presupposition of events occurring or states holding, such as speculation and uncertainty. Examples of speculative expressions are 'speculated' and 'maybe'.

Following are examples of strong negative evaluative language from a corpus of Usetnet newsgroup messages:

(3a) I had in mind your facts, buddy, not hers.

(3b) Nice touch. "Alleges" whenever facts posted are not in your persona of what is "real".

One of our current efforts is to recognize such inflammatory language, to be able to automatically identify "flames" in Internet forums.

Following is an example of opinionated, editorial language, taken from an editorial in the Wall Street Journal:

(4) We stand in awe of the Woodstock generation's ability to be unceasingly fascinated by the subject of itself.

Sentences (5) and (6) illustrate the fact that sentences about speech events may be subjective or objective:

(5) Northwest Airlines settled the remaining lawsuits filed on behalf of 156 people killed in a 1987 crash, but claims against the jetliner's maker are being pursued, a federal judge said.

(6) "The cost of health care is eroding our standard of living and sapping industrial strength," complains Walter Maher, a Chrysler health-and-benefits specialist.

In (5), the material about lawsuits and claims is presented as factual information, and a federal judge is given as the source of information. In (6), in contrast, a complaint is presented. An NLP system performing information extraction on (6) should not treat the material in the quoted string as factual information, with the complainer as a source of information, whereas a corresponding treatment of sentence (5) would be appropriate.

Subjective sentences often contain individual expressions of subjectivity. Examples are 'fascinating' in (1), and 'eroding', 'sapping', and 'complains' in (6). The following paragraphs mention aspects of subjectivity expressions that are relevant for NLP applications.

First, although some expressions, such as '!', are subjective in all contexts, many, such as 'sapping' and 'eroding', may or may not be subjective, depending on the context in which they appear. A *potential subjective element* (*PSE*) is a linguistic element that may be used to express subjectivity. A *subjective element* is an instance of a potential subjective element, in a particular context, that is indeed subjective in that context (Wiebe, 1994).

Second, a subjective element expresses the subjectivity of a *source*, who may be the writer or someone mentioned in the text. For example, the source of 'fascinating' in (1) is the writer, while the source of the subjective elements in (6) is Maher. In addition, a subjective element has a *target*, i.e., what the subjectivity is about or directed toward. In (1), the target is a tale; in (6), the target of Maher's subjectivity is the cost of health care. These are examples of *object-centric subjectivity*, which is about an object mentioned in the text (other examples: 'I love this project'; 'The software is horrible'). Subjectivity may also be *addressee-oriented*, i.e., directed toward the listener or reader

[1] The term *subjectivity* is due to Ann Banfield (1982). For references to work on subjectivity, please see (Banfield, 1982; Fludernik, 1993; Wiebe, 1994; Stein and Wright, 1995).

(e.g., 'You are an idiot').

Third, there may be multiple subjective elements in a sentence, possibly of different types and attributed to different sources and targets. For example, in (4), subjectivity of the Woodstock generation is described (specifically, its fascination with itself). In addition, subjectivity of the writer is expressed (e.g., 'we stand in awe'). As described below, individual subjective elements were annotated as part of this work, refining previous work on sentence-level annotations.

Finally, of particular importance for this paper, PSEs may be complex expressions such as 'village idiot', 'powers that be', 'You' *NP*, and 'What a' *NP*. There is a great variety of such expressions, including many studied under the rubric of idioms (see, for example, (Nunberg et al., 1994)). This paper demonstrates a straightforward method for identifying some types of collocational PSEs.

3 Previous Work on Subjectivity Tagging

In previous work (Wiebe et al., 1999; Bruce and Wiebe, 1999), a corpus of sentences from the Wall Street Journal Treebank Corpus (Marcus et al., 1993) were manually annotated with subjectivity classifications by multiple judges. The judges were instructed to consider a sentence to be subjective if they perceived any significant expression of subjectivity (of any source) in the sentence, and to consider the sentence to be objective, otherwise. After two rounds of tagging, an average pairwise κ value of .7 was achieved on a test set. The EM learning algorithm was used to produce corrected tags representing the consensus opinions of the taggers (Goodman, 1974; Dawid and Skene, 1979). An automatic system to perform subjectivity tagging was developed using the new tags as training and testing data. In 10-fold cross validation experiments, a probabilistic classifier obtained an average accuracy on subjectivity tagging of 72.17%, more than 20 percentage points higher than a baseline accuracy obtained by always choosing the more frequent class. Five part-of-speech features, two lexical features, and a paragraph feature were used.

The above work demonstrated the feasibility of automatic subjectivity tagging by achieving significant improvements in accuracy using very simple features. The next step is to identify a richer stable of features. In (Wiebe, 2000), Lin's (1998) method for clustering words according to distributional similarity, seeded by a small amount of detailed manual annotation, was successfully used to automatically identify adjectival clues of subjectivity. There are two parameters of this process, neither of which was varied in (Wiebe, 2000): C, the cluster size considered, and FT, a filtering threshold, such that, if the seed word and the words in its cluster have, as a set, lower precision than the filtering threshold on the training data, the entire cluster, including the seed word, is filtered out. This process is adapted for use in the current paper, as described in section 6.

4 Training Data

The training data for this work consists of multiple corpora annotated at the expression level. In expression-level tagging, the judges first identify the sentences they believe are subjective. They next identify the subjective elements in the sentence, i.e., the expressions they feel are responsible for the subjective classification. For example (subjective elements are in parentheses):

They promised (yet) more for (really good stuff).

(Perhaps you'll forgive me) for reposting his response.

The main source of disagreement among the judges is subjective-element boundaries. For the current paper, boundaries between adjacent subjective elements are ignored.

The training data is the aggregation of three different corpora annotated at the expression level. *WSJ-SE* is a corpus of 1001 sentences of the Wall Street Journal Treebank Corpus (Marcus et al., 1993). Two judges were asked to identify the subjective elements in WSJ-SE. *NG-FE* is a corpus of 362 Usenet newsgroup messages, balanced among the categories alt, sci, comp, and rec in the Usenet hierarchy. Two judges were asked to identify the flame elements in NG-FE, where flame elements are the subset of subjective elements that are perceived to be inflammatory. Corpus *NG-SE* is a different set of Usenet newsgroup messages, containing 15413 words. A single judge performed subjective element annotations on this corpus. For corpora WSJ-SE and NG-FE, which were annotated by two judges, the union of the judges' annotations are used. All corpora were stemmed (Karp et al., 1992) and part-of-speech tagged (Brill, 1992).

5 Extracting Collocations

Much previous work in mining collocations from texts (e.g., (Smajda, 1993; Lin, 1999; Biber, 1993)) was directed at extending lexicographic resources to serve machine translation and word sense disambiguation. In this work, we mine data for potentially subjective collocations.

Collocations were mined from the three corpora based on the following method. First, all 1-grams, 2-grams, 3-grams, and 4-grams were extracted from the training data, and the precision of each was calculated. The precision of an n-gram is the number of subjective instances of that n-gram divided by the total number of instances of that n-gram. An instance of an n-gram is subjective if each word occurs in a subjective element. As mentioned above, boundaries between subjective elements are ignored, so there is no restriction that all words of the n-gram appear in a single subjective element.

Potentially subjective collocations were selected based on their precision, using two criteria. First, the precision of the n-gram must be at least 0.1. Second, the precision of the n-gram must be at least as good as the precisions of its constituents.

For example, let $(W1, W2)$ be a bi-gram consisting of consecutive words $W1$ and $W2$. $(W1,W2)$ is identified to be a potential subjective element if $precision(W1, W2) > .1$ and (where pc is $precision$):

$$pc(W1, W2) > max(pc(W1), pc(W2))$$

For tri-grams, we extend the second condition in the following way. Let $(W1, W2, W3)$ be a tri-gram consisting of consecutive words $W1$, $W2$, and $W3$. Then the condition is:

$$pc(W1, W2, W3) > max(pc(W1, W2), pc(W3)) \text{ or}$$

$$pc(W1, W2, W3) > max(pc(W1), pc(W2, W3))$$

4-grams were selected in the same manner as 3-grams, comparing the 4-gram with first the maximum of the precisions of word $W1$ and tri-gram ($W2$, $W3$, $W4$) and then with the maximum of the precisions of tri-gram ($W1$,$W2$,$W3$) and $W4$. The n-gram collocations identified as above will be called *fixed-n-grams*, i.e., they are fixed sequences of words of length n.

In this paper, a strict low-frequency requirement is used: a word must appear just once in the corpus, i.e., it must be unique in the corpus (future work will investigate tuning this requirement to the size of the corpus, or employing existing frequency lists of English words). Thus, for extracting and selecting the generalized collocations, we first find every word that appears just once in the corpus and replace it with 'UNIQUE'. In essence, we treat the set of single-instance words as a new, frequently-occurring word. The above method for extracting and selecting n-grams is then used to obtain the potentially subjective collocations with positions filled by UNIQUE. Hereafter we will refer to the collocations as *ugen-n-grams* (**u**nique **gen**eralized-n-grams).

6 Experiments

To test the potentially subjective n-grams extracted from WSJ-SE, NG-FE, and NG-SE using the automatic process outlined above, three datasets of the Wall Street Journal from the Treebank corpus were used as test data (WSJ-10, WSJ-22, and WSJ-33). Each consists of roughly 160,000 words, and they are distinct from the WSJ corpus in the training data (WSJ-SE).

An advantage of using the Wall Street Journal is that there are existing annotations to exploit. Specifically, there are articles explicitly identified to be *Editorials, Letters to the Editor, Viewpoints,* and *Arts & Leisure Reviews.* Together, we call these *opinion pieces*. We discovered, however, that some editorials are not marked as such. That is, they are written in the first person, and the purpose of the article is to present an argument, rather than cover a news story; however, there is no explicit indication that they are editorials. Because this data is being used for evaluation, we made a pass through this data, marking these additional editorials.

To fully understand the problem of recognizing opinion pieces, it is important to understand that opinion pieces contain objective sentences; likewise, non-opinion pieces contain a large number of sentences that are subjective. Editorials frequently include facts to support their arguments, and traditional news reports include reactions to the news event (van Dijk, 1988). For examples, the views of supporters of a bill might be presented, as well as the views of its critics. In addition, quoted-speech sentences in which individuals express their subjectivity are often included (Barzilay et al., 2000).

According to our annotations of WSJ-SE, in opinion pieces, 70% of sentences are subjective and 30% objective. Even more indicative of the difficulty of the problem, in non-opinion pieces, 44% of the sentences are subjective and only 56% are objective. In addition, the distribution of opinion and non-opinion pieces is highly skewed in favor of non-opinion pieces. These facts, coupled with the fact that we are searching for potential subjective elements (PSEs) which may also have objective uses (see section 2), means that the absolute precision figures presented in this paper are low. We use increases in precision over a baseline precision as evidence that we have found promising clues of subjectivity. The baseline for comparison is the number of word instances in opinion pieces, divided by the total number of word instances.

Table 1 shows the results of testing the fixed-n-gram and the ugen-n-gram patterns on the three WSJ corpora. The *freq* columns give total frequencies, and the *+prec* columns show the improvements in precision from the baseline. The baseline precisions are given in the bottom of the table. The precisions of the collocations are all best on WSJ-10. The precisions of the fixed-2-grams and fixed-3-grams show modest improvements over the baseline (the frequencies of the

	WSJ-10		WSJ-22		WSJ-33	
	freq	+prec	freq	+prec	freq	+prec
fixed-2-grams	2182	.07	2151	.04	2080	.04
ugen-2-grams	294	.24	286	.15	288	.14
fixed-3-grams	271	.09	230	.06	262	.06
ugen-3-grams	132	.27	153	.12	144	.13
fixed-4-grams	32	.05	27	.18	30	-0.03
ugen-4-grams	3	.83	15	.15	15	.27
baseline precision		.17		.11		.13

freq: Total frequency **+prec:** Increase in precision over baseline

Table 1: Frequencies and Increases in Precision of Collocations

fixed-2-grams are high). The performance of the ugen-n-grams are very interesting, showing large increases in precision.

These results show that the method used to extract and select potentially subjective n-grams is promising. However, two questions come to mind. First, how do these potentially subjective collocations compare to other types of features for recognizing subjectivity?

Second, how do the sets of instances extracted by the fixed-n-grams compare to the set of instances extracted by the ugen-n-grams? Are they relatively similar, or are the two applications of the method recognizing different PSEs?

To address the first question, we used adjective and verb subjectivity clues identified using distributional similarity as described above in section 3. As mentioned above, (Wiebe, 2000) showed success automatically identifying adjective PSEs using Lin's method, seeded by a small amount of detailed manual annotations. Desiring to move away from manually annotated data, for this paper the same process is used, but the seed words are all the adjectives (verbs) in the training data. In addition, in the current setting, there are no a priori values to use for parameters C (cluster size) and FT (filtering threshold), as there were in (Wiebe, 2000), and results vary with different parameter settings. Thus, a train-validate-test process is appropriate. In Table 2, the numbers given under, e.g., W9-10, are the results obtained when W9-10 is used as the test set. One of the other datasets, say W9-22, was used as the training set, meaning that all the adjectives (verbs) in that dataset are the seed words, and all filtering was performed using only that data. The seed-filtering process was repeated with different settings of C and FT, producing a different set of adjectives (verbs) for each setting. A third dataset, say W9-33, was used as a validation set, i.e., among all the sets of adjectives generated from the training set, those with good performance on the validation set were selected as the PSEs to test on the test set. A set was considered to have good performance on the validation set if its precision is at least .25 and its frequency is at least 100. Since this process is meant to be a method for mining existing document-level annotations for PSEs, the existing opinion-piece annotations were used for training and validation. Our manual opinion-piece annotations were used for testing, as they were for table 1.

When comparing the results in Tables 1 and 2, two very interesting points emerge. First, although ugen-n-grams have lower frequencies, in every dataset both ugen-2-grams and ugen-3-grams have higher precisions than either adjectives or verbs. Second, it is interesting to note that all of the features (fixed-n-grams, ugen-n-grams, adjectives, verbs, and unique uni-grams) do well on the WSJ-10 corpus, but worse on the other two corpora. This shows that the subjectivity in WSJ-33 and WSJ-22 is somehow harder to identify than that in WSJ-10. It also shows an important consistency of performance among subjectivity clues identified by different processes on different types of data.

To address the second question, i.e., how the sets of instances extracted by the ugen-n-grams compare to the sets of instances extracted by the fixed-n-grams, we examined their intersections. For each WSJ corpus, the set of fixed-2-gram instances was compared with the set of ugen-2-gram instances, the set of fixed-3-gram instances was compared to the set of ugen-3-gram instances, and similarly for 4-grams. Table 3 shows the results of this analysis. Contrary to what we expected, we found virtually no overlap between sets, indicating that completely different types of potentially subjective collocations are being identified. More work needs to be done investigating the actual collocation instances in each of these sets, but a surface scan of the resulting collocations suggests that the fixed-n-grams which we are finding are those composed of more frequent words, while the ugen-n-grams seem to be collocations which are more like extraction patterns (Riloff and Jones, 1999).

For example, following are the fixed-3-gram collo-

	WSJ-10		WSJ-22		WSJ-33	
	freq	+prec	freq	+prec	freq	+prec
adjectives	373	.21	1340	.10	2137	.09
verbs	721	.16	1436	.07	3139	.07
unique 1-grams	6065	.10	5441	.06	6048	.06
baseline precision		.17		.11		.13
freq: Total frequency **+prec:** Increase in precision over baseline						

Table 2: Frequencies and Increases in Precision of Uni-Gram Features

WSJ-10	2grams	3grams	4grams
intersecting instances	4	2	0
%overlap	0.0016	0.0049	0
WSJ-22			
intersecting instances	4	0	0
%overlap	0.0016	0	0
WSJ-33			
intersecting instances	0	0	0
%overlap	0	0	0

Table 3: Overlap between fixed-i-grams and ugen-i-grams, i=2,3,4

cations that appear in opinion pieces in at least two of the WSJ datasets: as he be, be in the, be the case, have to pay, he be a, in the air, in the middle, it be time, it should be, of the century, rest of the, rest of us, seem to be, should have be, some of us, the country be, the difference between, the kind of, the middle of, the need to, the other hand, the quality of, the rest of, to be the, to do so, to say about.

These words tend to be more common; it is their use together in a collocation that makes them subjective.

Following are a sample of the ugen-3-grams that appear in opinions in at least two of the WSJ datasets ('U' is used for 'UNIQUE'). Note that some ugen-n-grams have more than one part generalized by UNIQUE. *U-adj and-CC U-adj*: charming and cheerful; pointless and unattainable. *U-adj to-TO U-verb*: wise to temper, allergic to sulfa; *U-noun and-CC U-verb*: condescension and misplace, adjudication and rulemaking; *U-noun and-CC a-DT*: remembrance and a, loyalist and a; *U-noun be-verb U-adj*: Dunaway be superb, idealism be hopeless; *U-noun be-verb a-DT*: caricature be a, Swartz be a; *U-noun of-IN U-noun*: separation of church, adventure of Beaumarchais; *U-noun on-IN the-DT*: team-work on the, fire-fighter on the; *U-verb and-CC U-verb*: bleat and bore, shimmy and rattle; *U-verb the-DT U-noun*: tolerate the sappy, impoverish the populace; *U-verb to-TO a-DT:* trek to a, entrust to a; *and-CC U-adverb U-adj:* and dazzlingly virtuoso, and wholly supportive; *a-DT U-noun for-IN:* a cure for, a propensity for; *be-verb an-DT U-noun:* be an agglutination, be an archconservative; *it-pronoun be-verb U-adverb:* it be sooo, it be best; *the-DT U-noun on-IN:* the workplace on, the plug on; *to-TO U-verb to-TO:* to entrust to, to defect to; *to-TO their-pronoun U-noun:* to their bosom, to their waist; *with-IN a-DT U-noun:* with a vengeance, with a denunciation.

7 Recognizing Opinionated Documents

This section shows that adding the fixed-n-grams and ugen-n-grams into the feature mix gives us more knowledge for recognizing opinionated documents.

Consider Table 4. The rows labeled *in opinions* show the number of features in opinions divided by the total number of words in opinions. Similarly for the rows labeled *out opinions*. These figures are given for various combinations of the features already presented in this paper. As unique words and then n-grams are added, the ratio between the feature counts *in opinions* and *out opinions* remains approximately the same, even as the counts go up.

At the bottom of Table 4 is another feature, average maximum density. Maximum density is defined to be the maximum number of features that can be found in a window of 11 words in a given document. We calculated the maximum density for each document using all features, and then averaged the maximum densities over opinion and non-opinion pieces.

Using the various feature sets, we ran simple linear regressions on the best dataset, WSJ-10, to see how they would do in classifying opinion pieces. The results can be found in Table 5. Column 1 shows the percentage of documents correctly classified for each set of features. With the skewed distribution, the re-

	WSJ-10	WSJ-22	WSJ-33
adjs, verbs			
in opinions	0.014	0.030	0.053
out opinions	0.005	0.016	0.030
adjs, verbs, uniques			
in opinions	0.074	0.082	0.108
out opinions	0.038	0.048	0.066
adjs, verbs, uniques, ngrams			
in opinions	0.102	0.110	0.133
out opinions	0.054	0.064	0.082
average max density			
in opinions	5.2	5.5	6.4
out opinions	2.9	3.4	3.4

Table 4: Counts and Densities of Feature Combinations

sults are not dramatic. However, we do find promise when looking at the true positives and false positives. As features are added, the number of true positives doubles while the number of false positives remains fixed.

8 Conclusions and Future Work

Knowledge of subjective language would be useful in many NLP applications, from information extraction to review mining and flame recognition. Collocations are an important type of subjectivity clue. This paper has demonstrated a straightforward method for learning certain kinds of potentially subjective collocations from corpora. The method is first applied directly to the words in the text. It is then applied again to the same data, but with all words that appear just once replaced by a single word (UNIQUE). The two applications yield different results, and both show increased precision. They also show consistency in performance with subjectivity clues identified in previous work and, together with those clues, show promise for recognizing opinionated documents. In addition to classifying editorials in news reports that are not marked as such, this task is important for filtering the results of search engines.

Future work on learning potentially subjective collocations from text include methods such as (Lin, 1999) for recognizing non-compositional phrases, and mutual bootstrapping (Riloff and Jones, 1999) for alternatively recognizing fixed sequences and subjective uni-grams. For example, ‘What a’ can be used as an extraction pattern for subjective NPs such as ‘idiot’, ‘fool’, and other insults.

Unique instances and collocations are two of the most challenging aspects of recognizing subjectivity. This paper makes some headway in exploiting both.

References

C. Aone, M. Ramos-Santacruz, and W. Niehaus. 2000. Assentor: An nlp-based solution to e-mail monitoring. In *Proc. IAAI-2000*, pages 945–950.

A. Banfield. 1982. *Unspeakable Sentences*. Routledge and Kegan Paul, Boston.

R. Barzilay, M. Collins, J. Hirschberg, and S. Whittaker. 2000. The rules behind roles: Identifying speaker role in radio broadcasts. In *Proc. AAAI*.

D. Biber. 1993. Co-occurrrence patterns among collocations: A tool for corpus-based lexical knowledge acquisition. *Computational Linguistics*, 19(3):531–538.

E. Brill. 1992. A simple rule-based part of speech tagger. In *Proc. of the 3rd Conference on Applied Natural Language Processing (ANLP-92)*, pages 152–155.

R. Bruce and J. Wiebe. 1999. Recognizing subjectivity: A case study of manual tagging. *Natural Language Engineering*, 5(2).

A. P. Dawid and A. M. Skene. 1979. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Applied Statistics*, 28:20–28.

M. Fludernik. 1993. *The Fictions of Language and the Languages of Fiction*. Routledge, London.

L. Goodman. 1974. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61:2:215–231.

E. Hovy. 1987. *Generating Natural Language under Pragmatic Constraints*. Ph.D. thesis, Yale University.

E. Hovy. 1998. Combining and standardizing large-scale practical ontologies for machine translation and other uses. In *In Proc. 1st International conference on language resources and evaluation (LREC)*.

	%classified	true positives	false positives
adjs, verbs	0.896	5	4
ngrams	0.899	5	3
adjs, verbs, ngrams	0.909	9	4
adjs, verbs, uniques, ngrams	0.909	9	4
all features (+max density)	0.912	11	4
baseline: 274/307 = .892			

Table 5: Classification of opinion and non-opinion pieces in WSJ-10

D. Karp, Y. Schabes, M. Zaidel, and D. Egedi. 1992. A freely available wide coverage morphological analyzer for English. In *Proc. of the 14th International Conference on Computational Linguistics (COLING-92)*.

D. Kaufer. 2000. *Flaming: A White Paper*. www.eudora.com.

B. Kessler, G. Nunberg, and H. Schutze. 1997. Automatic detection of text genre. In *Proc. ACL-EACL-97*.

D. Lin. 1998. Automatic retrieval and clustering of similar words. In *Proc. COLING-ACL '98*, pages 768–773.

D. Lin. 1999. Automatic identification of non-compositional phrases. In *Proc. ACL-99*, pages 317–324.

M. Marcus, Santorini, B., and M. Marcinkiewicz. 1993. Building a large annotated corpus of English: The penn treebank. *Computational Linguistics*, 19(2):313–330.

G. Nunberg, I. Sag, and T. Wasow. 1994. Idioms. *Language*, 70:491–538.

P. Procter. 1978. *Longman Dictionary of Contemporary English*. Addison Wesley Longman.

E. Riloff and R. Jones. 1999. Learning dictionaries for information extraction by multi-level bootstrapping. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI99)*, pages 1044–1049.

W. Sack. 1995. Representing and recognizing point of view. In *Proc. AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*.

F. Smajda. 1993. Retrieving collocations from text: Xtract. *Computational Linguistics*, 19.

E. Spertus. 1997. Smokey: Automatic recognition of hostile messages. In *Proc. IAAI*.

D. Stein and S. Wright, editors. 1995. *Subjectivity and Subjectivisation*. Cambridge University Press, Cambridge.

L. Terveen, W. Hill, B. Amento, D. McDonald, and J. Creter. 1997. Building task-specific interfaces to high volume conversational data. In *Proc. CHI 97*, pages 226–233.

S. Teufel and M. Moens. 2000. What's yours and what's mine: Determining intellectual attribution in scientific texts. In *Proc. Joint SIGDAT Converence on EMNLP and VLC*.

T.A. van Dijk. 1988. *News as Discourse*. Lawrence Erlbaum, Hillsdale, NJ.

J. Wiebe, R. Bruce, and T. O'Hara. 1999. Development and use of a gold standard data set for subjectivity classifications. In *Proc. 37th Annual Meeting of the Assoc. for Computational Linguistics (ACL-99)*, pages 246–253, University of Maryland, June. ACL.

J. Wiebe. 1994. Tracking point of view in narrative. *Computational Linguistics*, 20(2):233–287.

J. Wiebe. 2000. Learning subjective adjectives from corpora. In *17th National Conference on Artificial Intelligence (AAAI-2000)*.