Learning to Disambiguate Potentially Subjective Expressions

Janyce Wiebe and **Theresa Wilson**
Department of Computer Science and **Intelligent Systems Program**
University of Pittsburgh
Pittsburgh, PA, USA, 15260

Abstract

The goal of this work is recognizing opinionated and evaluative (*subjective*) language in text. The ability to recognize such language would be beneficial for many NLP applications such as question answering, information extraction, summarization, and genre detection. This paper focuses on disambiguating potentially subjective expressions in context, based on the density of other clues in the surrounding text.

1 Introduction

The goal of this work is to recognize opinionated and evaluative (*subjective*) language in text (Banfield, 1982). The ability to recognize subjective language would be beneficial for NLP applications such as question answering, information extraction, and genre detection.

Recent work by Wiebe and colleagues (2001b; 2001a; 2000) focused on learning potentially subjective expressions from corpora. This paper focuses on the mutual disambiguation of such features in context. Many natural language expressions have both subjective and objective usages, so a problem for recognizing subjective language is determining when instances of expressions are indeed subjective in the context in which they appear. We have discovered that the density of other potential clues in the surrounding context is a strong influence: if a sufficient number of other clues are nearby, a clue is more likely to be subjective than if there are not. There are two parameters to this process, corresponding to "sufficient number" and "nearby." Values for these parameters are chosen independently using manual annotations of subjective expressions in a mixture of Wall Street Journal (WSJ) and newsgroup (NG) data (Section 5).

The selected density features are evaluated with respect to document-level classes in WSJ data (Section 6). All of the parameters chosen using the manual annotations result in increases in precision over baseline in the test data, and the majority of the increases are large.

The document-level classes are identified by the WSJ itself: *Editorials, Letters to the Editor*, *Arts & Leisure Reviews*, and *Viewpoints*; together, we call these *opinion pieces.* With this data, we are not restricted to the confines of corpora manually annotated in detail, which is necessarily small. To assess the subjectivity of the sentences being recognized, Section 7 presents the results of an annotation study in which sentences identified automatically using density features are manually annotated by two judges. Agreement is high for sentences classified with certainty, and most of the sentences are classified as subjective by both judges, or are near sentences that are.

This work is also an interesting case study of using data annotated at different levels, and exploiting existing document-level annotations to learn linguistic knowledge.

2 Subjectivity

Subjectivity in natural language refers to aspects of language used to express emotion, evaluation, opinion and speculation. In this work, we adopt the annotation scheme of (Wiebe et al., 1999). Under that scheme, a sentence is subjective if it contains a significant expression of emotion, evaluation, opinion, or speculation, attributed to either the writer or someone mentioned in the text. Otherwise, the sentence is objective.

In (Wiebe et al., 1999), multiple judges annotated a corpus with subjective/objective classifications, rating the certainty of their answers on a scale from 0 to 3. For the 85% of the corpus for which the certainty ratings of the judges was 2 or 3, the average pairwise Kappa value was

0.80. Thus, when the annotators are certain of their answers, which they are for the majority of sentences, their agreement is high. Wiebe et al. (1999) developed a classifier to perform subjectivity tagging using this data, with good results in 10-fold cross validation experiments (an average accuracy 20 percentage points higher than baseline for all sentences and 30 percentage points higher on the sentences for which the annotators' certainty ratings were 2 or 3). In the current paper, as described below, this data is further annotated at the expression level and used for training data to choose density parameters.

Table 1 shows examples of subjective and objective sentences from the annotation study presented in Section 7. Sentences classified by both judges as objective are marked "oo" and those classified by both judges as subjective are marked "ss".

Subjectivity analysis could be exploited in many NLP applications, recognizing inflammatory messages (Spertus, 1997), genre detection and document routing (Kessler et al., 1997), intellectual attribution in text (Teufel and Moens, 2000), generation and style (Hovy, 1987), question answering from multiple perspectives, and any other application that would benefit from knowledge of how opinionated language is, and whether or not the writer purports to objectively present factual material. An information extraction or summarization system, for example, would benefit from distinguishing sentences intended to present facts from those intended to present opinions, since many such systems are meant to extract only facts.

One aspect of subjectivity is highlighted in this paper. Although some expressions, such as *!*, are subjective in all contexts, many may or may not be subjective, depending on the context in which they appear. A *potential subjective element* (**PSE**) is a linguistic word or expression 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). This paper focuses on identifying PSE instances that are subjective elements.

3 Data

We use training data from (Wiebe et al., 1999; Wiebe et al., 2001b; Wiebe et al., 2001a) consisting of corpora annotated at the expression level. In expression-level subjectivity 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, an annotator marked two subjective elements in the following sentence (indicated with parentheses): *They paid (yet) more for (really good stuff).*

Two WSJ datasets, 500 sentences each, were annotated by two judges, resulting in four sets of annotations (*SE1, SE2, SE3, SE4*). In addition, a newsgroup dataset of 1,132 sentences was annotated by one judge (*SE5*).

Document-level opinion-piece data is used as test data to evaluate the density features (in Section 6). Recall that the class *opinion piece* is the union of *Editorial, Letter to the Editor, Arts & Leisure Review*, and *Viewpoint* in the WSJ. An inspection of some data revealed that some editorials are not marked as such. Thus, the opinion-piece data used for evaluation in this work has been manually refined. The annotation instructions are simply to identify any additional opinion pieces that are not marked as such. To test the reliability of this annotation, two judges independently annotated two editions of the WSJ, each approximately 160K words. This is an "annotation lite" task: with no training, the annotators achieved Kappa values of 0.94 and 0.95, and spent an average of three hours per WSJ edition.

Two datasets (*OP1* and *OP2*) of four WSJ editions each were manually annotated as described in the previous paragraph. *OP1* has a total of 1,232 articles and 640,975 words, and *OP2* has a total of 1,222 documents and 629,690 words. All instances in *OP1* and *OP2* of the PSEs described in Section 4 were identified. All training to define the PSE instances in $OP1$ was performed on data separate from $OP1$, and all training to define the PSE instances in $OP2$ was performed on data separate from $OP2$.

Note that opinion-piece data, used as test data in our evaluations, is noisy. Although the ratio of subjective to objective sentences is higher in opinion-piece documents, there are

(1)	The outburst of shooting came nearly two weeks after clashes between Moslem worshippers and Somali soldiers.	**oo**
(2.a)	But now the refugees are streaming across the border and alarming the world.	**ss**
(2.b)	In the middle of the crisis, Erich Honecker was hospitalized with a gall stone operation.	**oo**
(2.c)	It is becoming more and more obvious that his gallstone-age communism is dying with him: ...	**ss**
(3.a)	Not brilliantly, because, after all, this was a performer who was collecting paychecks from lounges at Hiltons and Holiday Inns, but creditably and with the air of someone for whom "Ten Cents a Dance" was more than a bit autobiographical.	**ss**
(3.b)	"It was an exercise of blending Michelle's singing with Susie's singing," explained Ms. Stevens.	**oo**
(4)	Enlisted men and lower-grade officers were meat thrown into a grinder.	**ss**
(5)	"If you believe in God and you believe in miracles, there's nothing particularly crazy about that."	**ss**

Table 1: Examples from the annotation study in Section 7.

objective sentences in opinion-piece documents, and subjective sentences in the other kinds of documents.

4 PSEs Used

We use PSE features automatically learned from corpora. The first is a word appearing just once in the corpus (i.e., a *unique* word). Interestingly, the set of all unique words in a corpus is a high frequency set, with higher than baseline precision (Wiebe et al., 2001a). This feature does not require training.

The next two types of PSE are adjectives and verbs identified using the results of a method for clustering words according to distributional similarity (Lin, 1998), as described in (Wiebe, 2000). Distributional similarity has been used to find similar words in text. The hypothesis behind its use in (Wiebe, 2000) was that words may be distributionally similar because they are both potentially subjective (e.g., *tragic*, *sad*, and *poignant* are identified from *bizarre*).

The remaining types of PSEs are collocations, learned from data that is manually annotated at the expression level with subjective elements. Roughly speaking, a collocation is judged to be a PSE if its precision is greater than the maximum precision of its constituents (Wiebe et al., 2001a).

Fixed-n-grams are sequences of n word|part-of-speech pairs. Examples from test data $OP1$ are: *a sort of, as he be, be it that, have to pay, he be a, it be time, it should be, of the century, rest of us, seem to be, the kind of, the middle of, the other hand, the quality of, to do so, to say about.*

In *ugen-n-grams*, one or more of the words is U, which matches any word that is unique in the test data. Two examples are (highly|*adverb* U|*adj*) and (U|*adj* to|*prep* U|*verb*). Instances in $OP1$ matching the first include *highly unsatisfactory, highly unorthodox, highly talented, highly conjectural, highly erotic.* Instances in $OP1$ matching the second include: *impervious to reason, strange to celebrate, wise to temper.*

	Freq	Prec	%incPrec
unique words	8288	.32	100%
adjectives	4610	.34	113%
verbs	8862	.25	56%
fixed-2-grams	7584	.22	38%
fixed-3-grams	908	.23	44%
fixed-4-grams	61	.26	63%
ugen-2-grams	407	.43	169%
ugen-3-grams	203	.42	163%
ugen-4-grams	15	.47	194%
baseline	640975	.16	

Table 2: Results for PSEs in test data $OP1$

Table 2 gives results for the PSEs described above on test data $OP1$ (all training was done on separate data). The precision of a set S with respect to opinion pieces is the proportion of members of S that appear in opinion pieces. The baseline precision of .16 appearing at the bottom of the table is the proportion of all words in the corpus that appear in opinion pieces. For each type of feature, Table 2 gives frequencies (in column *Freq*), precisions (in column *Prec*), and percentage increases in precision over baseline (column *%incPrec*). For example, the first row gives results for the set of unique words. There are 8,288 unique words in *OP1*. The precision of that set is .32, which is a 100% improvement over the baseline precision of .16.

The baseline precision in Table 2 is low be-

0. $PSEs$ = all adjs, verbs, modals, nouns, and adverbs that appear at least once in an SE (except *not*, *will*, *be*, *have*).
1. $PSEinsts$ = the set of all instances of $PSEs$
2. $HiDensity = \{\}$
3. For P in $PSEinsts$:
    4. leftWin(P) = the W words before P
    5. rightWin(P) = the W words after P
    6. density(P) = # of SEs whose first or last word is in leftWin(P) or rightWin(P)
    7. if density(P) $\geq T$:
    $HiDensity = HiDensity \cup \{P\}$

8. $\text{prec}(PSEinsts) = \frac{\#\text{ of } PSEinsts \text{ in } SEs}{|PSEinsts|}$

9. $\text{prec}(HiDensity) = \frac{\#\text{ of } HiDensity \text{ in } SEs}{|HiDensity|}$

Figure 1: Algorithm for calculating density in subjective element (SE) data

cause the distribution is highly skewed in favor of non-opinion pieces.

5 Choosing Density Parameters from Subjective Element Data

In (Wiebe, 1994), whether a PSE is interpreted to be subjective depends, in part, on how subjective the surrounding context is. We explore this idea in the current work, assessing whether PSEs are more likely to be subjective if they are surrounded by subjective elements. In particular, we experiment with a density feature to decide whether or not a PSE instance is subjective: if a sufficient number of subjective elements are nearby, then the PSE instance is considered to be subjective; otherwise, it is discarded. The density parameters are a window size W and a frequency threshold T.

In this section, we explore density in the manually-annotated subjective-element (SE) data, and choose density parameters for later use in automatic disambiguation in separate test data (in Section 6). The process for calculating density in the subjective-element data is given in Figure 1. The PSEs are defined to be all adjectives, verbs, modals, nouns, and adverbs that appear at least once in a subjective element, with the exception of some stop words (line 0 of Figure 1). Note that these PSEs depend only on the subjective-element manual annotations, not on the automatically identified features used elsewhere in the paper, nor on the document-level opinion-piece classes. $PSEinsts$ is the set of PSE instances to be disambiguated (line 1). *HiDensity* (initialized on line 2) will be the subset of $PSEinsts$ that are retained. In the loop, the density of each PSE instance P is calculated, which is the number of subjective elements that begin or end in the W words preceding or following P (line 6). P is retained if its density is at least T (line 7).

The precision of a set S with respect to subjective-element classifications is the number of members of S that appear in subjective elements over the total number of members of S. Lines 8-9 assess the precision of the original ($PSEinsts$) and new ($HiDensity$) sets of PSE instances. If prec($HiDensity$) is greater than prec($PSEinsts$), then there is evidence that the number of subjective elements near a PSE instance is related to its subjectivity in context.

The process in Figure 1 was repeated for different parameter settings (T in $[1, 2, 4, \ldots, 48]$ and W in $[1, 10, 20, \ldots, 490]$) on each of the five subjective-element datasets. To find good parameter settings, the results for each dataset were sorted into 5-point precision intervals, and then sorted by frequency within each interval. Information for the top three precision intervals for each dataset are shown in Table 3, specifically the parameter values (i.e., T and W) and the frequency and precision of the most frequent result in each interval. The intervals are in the rows labeled "Range". For example, the top three precision intervals for $SE1$ are .77-.82, .82-.87, and .87-.92 (no parameter values yield higher precision than .92).

The top of Table 3 gives baseline frequencies and precisions, which are $|PSEinsts|$ and prec($PSEinsts$), respectively, in line 8 of Figure 1.

The parameter values exhibit a range of frequencies and precisions, with the expected tradeoff between precision and frequency. We choose the following parameters to test in Section 6 below: for each dataset (e.g., $SE1$), for each precision interval whose lower bound is at least 10 percentage points higher than the baseline for that dataset, the top two T,W pairs yielding the highest frequencies in that interval are chosen. Among the five datasets, a total of 45 parameter pairs were selected.

	SE1	SE2	SE3	SE3	SE5
freq	1566	1245	1167	1108	3303
prec	.49	.47	.41	.36	.51
Range	.87-.92	.95-1.0	.95-1.0	.95-1.0	.95-1.0
T,W	10,20	12,50	20,50	14,100	10,10
freq	76	12	1	1	3
prec	.89	1.0	1.0	1.0	1.0
Range	.82-.87	.90-.95	.73-.78	.51-.56	.67-.72
T,W	6,10	12,60	46,190	22,370	26,90
freq	63	22	53	221	664
prec	.84	.91	.78	.51	.67
Range	.77-.82	.84-.89	.66-.71	.46-.51	.63-.67
T,W	12,40	12,80	18,60	16,310	8,30
freq	292	42	53	358	1504
prec	.78	.88	.68	.47	.63

Table 3: Most frequent entry in the top 3 precision intervals for each subjective element dataset

0. $PSEinsts$ = the set of instances in the test data of all PSEs described in Section 4
1. $HiDensity = \{\}$
2. For P in $PSEinsts$:
   3. leftWin(P) = the W words before P
   4. rightWin(P) = the W words after P
   5. density(P) = # of $PSEinsts$ whose first or last word is in leftWin(P) or rightWin(P)
   6. if density(P) $\geq$ T: $HiDensity = HiDensity \cup \{P\}$
7. $\text{prec}(PSEinsts) = \frac{\#\text{ of } PSEinsts \text{ in } OPs}{|PSEinsts|}$
8. $\text{prec}(HiDensity) = \frac{\#\text{ of } HiDensity \text{ in } OPs}{|HiDensity|}$

Figure 2: Algorithm for calculating density in opinion piece (OP) data

6 Density for Disambiguation

In this section, density is exploited as an informative feature for PSE disambiguation. The process is shown in Figure 2. There are only two differences between the algorithms in Figures 1 and 2. First, in Figure 1, density was defined in terms of the number of subjective elements nearby. However, subjective-element annotations are not available in test data. In Figure 2, density is defined in terms of the number of other PSE instances nearby, where $PSEinsts$ consists of all instances of the automatically identified PSEs described in Section 4 and for which results are given in Table 2.

Second, in Figure 2, we assess precision with respect to the document-level classes: the precision of a set is now the number of set members appearing in documents that are classified as opinion pieces divided by the cardinality of the set (see lines 7-8 of Figure 2).

The test data is corpus $OP1$.

An interesting question arose when defining the PSE instances: what should be done with words that are identified to be PSEs (or parts of PSEs) according to multiple criteria? For example, *sunny*, *radiant*, and *exhilarating* are all unique in corpus $OP1$, and are all members of the adjective PSE feature defined for testing on $OP1$. Collocations add additional complexity. For example, consider the sequence *and splendidly*, which appears in the test data. The sequence *and splendidly* matches the ugen-2-gram (and|*conj* U|*adj*), and the word *splendidly* is unique (all instances of ugen-n-grams result in at least two matches: the ugen-n-gram, and a unique). In addition, more than one n-gram feature may be matched by a sequence. For example, *is it that* matches three fixed-n-gram features: *is it*, *is it that*, and *it that*.

In the current experiments, the more PSEs a word matches, the more weight it is given. The hypothesis behind this treatment is that additional matches represent additional evidence that a PSE instance is subjective. This hypothesis is realized as follows: each match of each member of each type of PSE is considered to be a PSE instance. Thus, among them, there are 11 members in $PSEinsts$ for the 5 phrases *sunny*, *radiant*, *exhilarating*, *and splendidly*, and *is it that*, one for each of the matches mentioned above.

The process in Figure 2 was performed with the 45 parameter-pair values (T and W) chosen from the subjective-element data as described in Section 5. Table 4 shows results for a subset of the 45 parameters, namely the most frequent parameter pair chosen from the top six precision intervals for each training set. The bottom of the table gives a baseline frequency and precision in $OP1$, defined as $|PSEinsts|$ and prec($PSEinsts$), respectively, on line 7 of Figure 2.

As can be seen, the density features result in substantial increases in precision. Among all 45 parameter pairs, the minimum percentage increase over baseline is 21%. 24% of the 45 pa-

	SE1	SE2	SE3	SE4	SE5
T,W	10,20	12,50	20,50	14,100	10,10
freq	237	3176	170	10510	8
prec	.87	.72	.97	.57	1.0
T,W	6,10	12,60	46,190	22,370	26,90
freq	459	5289	1323	21916	787
prec	.68	.68	.95	.37	.92
T,W	12,40	12,80	18,60	16,310	8,30
freq	1398	9662	906	24454	3239
prec	.79	.58	.87	.34	.67
T,W	12,50	10,70	14,50		
freq	3176	10995	1581		
prec	.72	.55	.81		
T,W	20,110	14,110	1,10		
freq	5330	12206	21221		
prec	.73	.53	.34		
T,W	6,40	12,100			
freq	11426	13637			
prec	.50	.50			
PSE Baseline: Freq=30938, Prec=.28					

Table 4: Results for high-density PSEs in test data $OP1$ using parameters chosen from subjective-element data

rameter pairs yield increases of 200% or more; 38% yield increases between 100% and 199%, and 38% yield increases between 21%-99%.

Notice that, except for one blip ($T, W = 6, 10$ under $SE1$), the precisions decrease and the frequencies increase as we go down each column in Table 4. The same pattern can be observed in the full table with all 45 parameter pairs (not included due to space). **But the parameter pairs are ordered in Table 4 based on performance in the manually-annotated subjective-element data,** not based on performance in the test data $OP1$. For example, the entry in the first row, first column ($T, W = 10, 20$) is the parameter pair giving the highest frequency in the top precision interval of $SE1$ (frequency and precision in $SE1$, using the process of Figure 1). Thus, the relative precisions and frequencies of the parameter pairs are carried over from the training to the test data.

7 Sentence Annotation

To assess the subjectivity of sentences with high-density PSEs, we extracted the sentences in corpus $OP2$ that contain at least one high-density PSE, and manually annotated them. We chose the density parameter pair T,W=12,30, based on its precision and frequency in $OP1$. This parameter setting yields results that are relatively high precision and low frequency. We chose a low-frequency setting to make the annotation study feasible.

	S	O	U
S	98	2	3
O	2	14	0
U	2	11	1

Table 5: Sentence annotation contingency table; judge 1's counts are in the rows and judge 2's counts are in the columns.

133 sentences were so identified. They are referred to below as the *system-identified* sentences.

The extracted sentences were independently annotated by two judges. One is a co-author of this paper (judge 1), and the other has performed subjectivity annotation before, but is not otherwise involved in this research (judge 2). Sentences were annotated according to the coding instructions of (Wiebe et al., 1999), which, recall, are to classify a sentence as subjective if there is a significant expression of subjectivity in the sentence, of either the writer or someone mentioned in the text. In addition to the subjective and objective classes, a judge could tag a sentence "uncertain" if he or she is unsure of his or her rating.

An equal number (133) of other sentences were randomly selected from the corpus to serve as controls. The 133 system-identified sentences and the 133 control sentences were randomly mixed together. The judges were asked to annotate all 266 sentences, not knowing which are system-identified and which are control. Each sentence was presented with the sentence that precedes and the sentence that follows it in the corpus, to provide some context for interpretation.

Judge 1 classified 103 of the system-identified sentences as subjective; 16 as objective; and 14 as uncertain. Judge 2 classified 102 of the system-identified sentences as subjective; 27 as objective; and 4 as uncertain. The contingency table is given in Table 5.

For most of the sentences (116 out of 133, or 87% of the corpus), neither judge rated the sentence as uncertain. The agreement between judges on those sentences is very high: the

Kappa value is 0.86. With all sentences included, the Kappa value is 0.60. Thus, most of the disagreements involve sentences tagged as uncertain.

The current paper is concerned with whether high-density PSEs are indicative of subjective text. An examination of the data from this perspective is illuminating.

For 98 of the sentences (set *SS*), judges 1 and 2 tagged the sentence as subjective. Among the other 35 sentences (those tagged objective, those upon which the judges disagree, etc), 20 (set *inBlock*) appear in a block of contiguous system-identified sentences that includes a member of *SS*. For example, in Table 1, (2.a) and (2.c) are in *SS* while (2.b) is in *inBlock*, and (3.a) is in *SS* while (3.b) is in *inBlock*. Thus, fully 89% of all sentences are either in *SS* or *inBlock*. Among the 15 other sentences, 6 are adjacent to subjective sentences that were not identified by our system (so were not annotated by the judges). All contain significant expressions of subjectivity of the writer or someone mentioned in the text, the criterion used in this work for classifying a sentence as subjective. Thus, 93% of the sentences containing high-density PSEs are subjective or are near subjective sentences.

8 Conclusions and Future Work

This paper investigates a contextual feature for recognizing subjectivity, which identifies clusters of potentially subjective expressions (*PSEs*). This density feature involves two parameters. We select parameter values using training data manually annotated at the expression level, and then test them on data annotated at the document level for opinion pieces. The PSEs in the training data are defined in terms of the manual annotations, while the PSEs in the test data are automatically identified from text. All of the selected parameters lead to increases in precision on the test data, the majority leading to increases over 100%. The large differences between training and testing suggest that our results are not brittle.

Using a density feature selected from a training set, sentences containing high-density PSEs were extracted from a separate test set, and manually annotated by two judges. Fully 93% are subjective sentences or are near subjective sentences.

There are many avenues for future work. Our immediate plans are to apply the system to large amounts of data, and then apply information extraction and bootstrapping techniques (Riloff and Jones, 1999) to identify subjective language that the system does not yet know. In addition, it would be illuminating to apply our system to data annotated with discourse trees (Carlson et al., 2001).

References

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

L. Carlson, D. Marcu, and M. E. Okurowski. 2001. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *2nd SIGDIAL Workshop on Discourse and Dialogue.*

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

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

D. Lin. 1998. Automatic retrieval and clustering of similar words. In *COLING-ACL-98.*

E. Riloff and R. Jones. 1999. Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping. In *AAAI-99.*

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

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

J. Wiebe, R. Bruce, and T. O'Hara. 1999. Development and use of a gold standard data set for subjectivity classifications. In *ACL-99.*

J. Wiebe, R. Bruce, M. Bell, M. Martin, and T. Wilson. 2001a. A corpus study of evaluative and speculative language. In *2nd SIGDIAL Workshop on Discourse and Dialogue.*

J. Wiebe, T. Wilson, and M. Bell. 2001b. Identifying collocations for recognizing opinions. In *ACL-01 Workshop on Collocation.*

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

J. Wiebe. 2000. Learning subjective adjectives from corpora. In *AAAI-00.*