Effects of Adjective Orientation and Gradability on Sentence Subjectivity

Vasileios Hatzivassiloglou
Department of Computer Science
Columbia University
New York, NY 10027
vh@cs.columbia.edu

Janyce M. Wiebe
Department of Computer Science
New Mexico State University
Las Cruces, NM 88003
wiebe@cs.nmsu.edu

Abstract

Subjectivity is a pragmatic, sentence-level feature that has important implications for text processing applications such as information extraction and information retrieval. We study the effects of dynamic adjectives, semantically oriented adjectives, and gradable adjectives on a simple subjectivity classifier, and establish that they are strong predictors of subjectivity. A novel trainable method that statistically combines two indicators of gradability is presented and evaluated, complementing existing automatic techniques for assigning orientation labels.

1 Introduction

In recent years, computational techniques for the determination of *lexical semantic* features have been proposed and evaluated. Such features include sense, register, domain specificity, pragmatic restrictions on usage, semantic markedness, and orientation, as well as automatically identified links between words (e.g., semantic relatedness, synonymy, antonymy, and meronymy). Automatically learning features of this type from large corpora allows the construction or augmentation of lexicons, and the assignment of semantic labels to words and phrases in running text. This information in turn can be used to help determine additional features at the lexical, clause, sentence, or document level.

This paper explores the benefits that some lexical features of adjectives offer for the prediction of a contextual sentence-level feature, *subjectivity*. Subjectivity in natural language refers to aspects of language used to express opinions and evaluations. The computational task addressed here is to distinguish sentences used to present opinions and other forms of subjectivity (*subjective sentences*, e.g., "At several different layers, it's a fascinating tale") from sentences used to objectively present factual information (*objective sentences*, e.g., "Bell industries Inc. increased its quarterly to 10 cents from 7 cents a share").

Much research in discourse processing has focused on task-oriented and instructional dialogs. The task addressed here comes to the fore in other genres, especially news reporting and Internet forums, in which opinions of various agents are expressed and where subjectivity judgements could help in recognizing inflammatory messages ("flames") and mining online sources for product reviews. Other tasks for which subjectivity recognition is potentially very useful include information extraction and information retrieval. Assigning subjectivity labels to documents or portions of documents is an example of *non-topical* characterization of information. Current information extraction and retrieval technology focuses almost exclusively on the subject matter of the documents. Yet, additional components of a document influence its relevance to particular users or tasks, including, for example, the evidential status of the material presented, and attitudes adopted in favor or against a particular person, event, or position (e.g., articles on a presidential campaign written to promote a specific candidate). In summarization, subjectivity judgments could be included in document profiles to augment automatically produced document summaries, and to help the user make relevance judgments when using a search engine.

Other work on subjectivity (Wiebe et al., 1999; Bruce and Wiebe, 2000) has established a positive and statistically significant correlation with the presence of adjectives. Since the mere presence of one or more adjectives is useful for predicting that a sentence is subjective, we investigate in this paper the effects of additional lexical semantic features of adjectives that can be automatically learned from corpora. We consider two such features: *semantic orientation*, which represents an evaluative characterization of a word's deviation from the norm for its semantic group (e.g., *beautiful* is positively oriented, as opposed to *ugly*); and *gradability*, which characterizes a word's ability to express a property in varying degrees. In the remainder of this paper, we first address adjective orientation in Section 2, summarizing a previously published method for automatically separating oriented adjectives into positive and negative classes. Then, Section 3 presents a novel method for learning gradable adjectives using a large corpus and a statistical feature combination model. In Section 4, we review earlier experiments on testing subjectivity using various features as predictors, and then present comparative analyses of the effects that orientation and gradability have on our ability to predict sentence subjectivity from adjectives. We show that both give us higher-quality features for recognizing subjective sentences, and conclude by discussing future extensions to this work.

α	Number of adjectives in test set ($|A_\alpha|$)	Number of links in test set ($|L_\alpha|$)	Average number of links for each adjective	Accuracy	Ratio of average group frequencies
2	730	2,568	7.04	78.08%	1.8699
3	516	2,159	8.37	82.56%	1.9235
4	369	1,742	9.44	87.26%	1.3486
5	236	1,238	10.49	92.37%	1.4040

Table 1: Evaluation of the adjective orientation classification and labeling methods (from (Hatzivassiloglou and McKeown, 1997)).

2 Semantic Orientation

The *semantic orientation* or *polarity* of a word indicates the direction the word deviates from the norm for its semantic group or *lexical field* (Lehrer, 1974). It is an evaluative characteristic (Battistella, 1990) of the meaning of the word which restricts its usage to appropriate pragmatic contexts. Words that encode a desirable state (e.g., *beautiful*, *unbiased*) have a positive orientation, while words that represent undesirable states have a negative orientation. Within the particular syntactic class of adjectives, orientation can be expressed as the ability of an adjective to ascribe in general a positive or negative quality to the modified item, making it better or worse than a similar unmodified item.

Most antonymous adjectives can be contrasted on the basis of orientation (e.g., *beautiful–ugly*); similarly, nearly synonymous terms are often distinguished by different orientations (e.g., *simple–simplistic*). While orientation applies to many adjectives, there are also those that have no orientation, typically as members of groups of complementary, qualitative terms (Lyons, 1977) (e.g., *domestic*, *medical*, or *red*). Since orientation is inherently connected with evaluative judgements, it appears to be a promising feature for predicting subjectivity.

Hatzivassiloglou and McKeown (1997) presented a method for automatically assigning a + or − orientation label to adjectives known to have some semantic orientation. Their method is based on information extracted from conjunctions between adjectives in a large corpus—because orientation constrains the use of the words in specific contexts (e.g., compare *corrupt and brutal* with **corrupt but brutal*), observed conjunctions of adjectives can be exploited to infer whether the conjoined words are of the same or different orientation. Using a shallow parser on a 21 million word corpus of Wall Street Journal articles, Hatzivassiloglou and McKeown developed and trained a log-linear statistical model that predicts whether any two adjectives have the same orientation with 82% accuracy. The predicted links of same or different orientation are automatically assigned a strength value (essentially, a confidence estimate) by the model, and induce a graph that can be partitioned with a clustering algorithm into components so that all words in the same component belong to the same orientation class.

Once the classes have been determined, frequency information is used to assign positive or negative labels to each class (there are slightly fewer positive terms, but with a significantly higher rate of occurrence than negative terms).

Hatzivassiloglou and McKeown applied their method to 1,336 (657 positive and 679 negative) adjectives which were all the oriented adjectives appearing in the corpus 20 times or more. Orientation labels were assigned to these adjectives by hand.[1] Subsequent validation of the initial selection and label assignment steps with independent human judges showed an agreement of 89% for the first step and 97% for the second step, establishing that orientation is a fairly objective semantic property. Because the accuracy of the method depends on the density of conjunctions per adjective, Hatzivassiloglou and McKeown tested separately their algorithm for adjectives appearing in at least 2, 3, 4, or 5 conjunctions in the corpus; their results are shown in Table 1.

In this paper, we use the model labels assigned by hand by Hatzivassiloglou and McKeown, and the labels automatically obtained by their method and reported in (Hatzivassiloglou and McKeown, 1997) with the following extension: An adjective that appears in k conjunctions will receive (possibly different) labels when analyzed together with all adjectives appearing in at least 2, 3, ..., k conjunctions; since performance generally increases with the number of conjunctions per adjective, we select as the orientation label the one assigned by the experiment using the highest applicable conjunctions threshold. Overall, we have labels for 730 adjectives[2], with a prediction accuracy of 81.51%.

3 Gradability

Gradability (or *grading*) (Sapir, 1944; Lyons, 1977, p. 271) is the semantic property that enables a word to participate in comparative constructs and to accept modifying expressions that act as intensifiers or diminishers. Gradable adjectives express properties in varying degrees of strength, relative to a norm either explicitly

---

[1]Some adjectives with unclear, ambiguous, or context-dependent orientation were excluded.

[2]Those appearing in the corpus in two conjunctions or more, since some conjunction data must be left out to train the link prediction algorithm.

	cold		civil	
	Unmodified by grading words	Modified by grading words	Unmodified by grading words	Modified by grading words
Uninflected	392	20	1,296	1
Inflected for degree	18	0	0	0

Table 2: Extracted values of gradability indicators, i.e., frequencies of the word with or without the specified inflection or modification, for two adjectives, one gradable (*cold*) and one primarily non-gradable (*civil*). The frequencies were computed from the 1987 Wall Street Journal corpus.

mentioned or implicitly supplied by the modified noun (for example, a *small planet* is usually much larger than a *large house*; cf. the distinction between *absolute* and *relative* adjectives made by Katz (1972, p. 254)). This relativism in the interpetation of gradable words indicates that gradability is likely to be a good predictor of subjectivity.

3.1 Indicators of gradability

Most gradable words appear at least several times in a large corpus either in forms inflected for degree (i.e., comparative and superlative), or in the context of grading modifiers such as *very*. However, non-gradable words may also occasionally appear in such contexts or forms under exceptional circumstances. For example, *very dead* can be used for emphasis, and *redder and redder* (as in "her face became redder and redder") can be used to indicate a progression of coloring. To distinguish between truly gradable adjectives and non-gradable adjectives in these exceptional contexts, we have developed a trainable log-linear statistical model that takes into account the number of times an adjective has been observed in a form or context indicating gradability relative to the number of times it has been seen in non-gradable contexts.

We use a shallow parser to retrieve from a large corpus tagged for part-of-speech with Church's PARTS tagger (Church, 1988) all adjectives and their modifiers. Although the most common use of an adverb modifying an adjective is to function as an intensifier or diminisher (Quirk et al., 1985, p. 445), adverbs can also add to the semantic content of the adjectival phrase instead of providing a grading effect (e.g., *immediately available*, *politically vulnerable*), or function as emphasizers, adding to the force of the base adjective and not to its degree (e.g., *virtually impossible*; compare **very impossible*). Therefore, we compiled by hand a list of 73 adverbs and noun phrases (such as *a little*, *exceedingly*, *somewhat*, and *very*) that are frequently used as grading modifiers. The number of times each adjective appears modified by a term form this list becomes a first indicator of gradability.

To detect inflected forms of adjectives (which, in English, always indicate gradability subject to the exceptions discussed earlier), we have implemented an automatic morphology analysis component. This program

recognizes several irregular forms (e.g., *good–better–best*) and strips the grading suffixes *-er* and *-est* from regularly inflected adjectives, producing a list of candidate base forms that if inflected would yield the original adjective (e.g., *bigger* produces three potential forms, *big*, *bigg*, and *bigge*). The frequency of these candidate base words is checked against the corpus, and the form with significantly higher frequency is selected. To guard against cases of base adjective forms that end in *-er* or *-est* (e.g., *silver*), the original word is also included among the candidates. The total number of times this procedure is successfully applied for each adjective becomes a second indicator of gradability.

3.2 Determining gradability

The presence or absence of each of the above two indicators results in a 2×2 frequency table for each adjective; examples for one gradable and one non-gradable adjective are given in Table 2. To convert these four numbers to a single decision on the gradability of the adjective, we use a log-linear model. Log-linear models (Santner and Duffy, 1989) construct a linear combination (weighted sum) of the predictor variables V_i,

$$\eta = \beta_0 + \sum_{i=1}^{n} \beta_i V_i$$

and relate it to the actual response R (in this case, 0 for non-gradable and 1 for gradable) via the so-called *logistic transformation*,

$$R = \frac{e^\eta}{1 + e^\eta}$$

Maximum likelihood estimates for the coefficients β_i are obtained from training samples for which the correct response R is known, using the *iterative reweighted nonlinear least squares* algorithm (Bates and Watts, 1988). This statistical model is particularly suited for modeling variables with a "yes"–"no" (binary) value, because, unlike linear models, it captures the dependency of R's variance on its mean (Santner and Duffy, 1989).

We normalize the counts for the two indicators of gradability, and the count of joint occurrences of both inflection and modification by grading modifiers, by dividing with the total frequency of the adjective in the corpus. In this manner, we obtain three real-valued predictors

Classified as gradable:

> **acceptable accurate afraid aware busy careful cautious cheap creative critical dangerous different disappointing** equal **fair familiar far favorable formal free frequent good grand inadequate intense interesting legitimate likely positive professional reasonable rich short-term significant slow solid sophisticated sound speculative thin tight tough uncertain widespread** worth

Classified as non-gradable:

> **additional alleged alternative annual antitrust automatic** certain **criminal cumulative daily deputy domestic** elderly **false financial first-quarter** full hefty **illegal institutional internal legislative long-distance military minimum monthly** moral **national official one-time other** outstanding **present prior prospective punitive regional scientific secondary sexual subsidiary taxable three-month three-year total tremendous two-year** unfair **unsolicited upper voluntary** white **wholesale world-wide** wrong

Figure 1: Automatically obtained classification of a sample of 100 adjectives as gradable or not. Correct decisions (according to the COBUILD-based reference model) are indicated in bold.

$V_i, i = 1, \ldots, 3$ for the log-linear model. We also consider a modified model, where any adjective for which any occurrence of simultaneous inflection and modification has been detected is automatically labeled gradable; the remaining two predictors are used to classify the adjectives that do not fulfill this condition. This modification is motivated by the fact that observing an adjective in such a context offers a very high likelihood of gradability.

3.3 Experimental results

We extracted from the 1987 Wall Street Journal corpus (21 million words) all adjectives with a frequency of 300 or more; this produced a collection of 496 words. Gradability labels specifying whether each word is gradable or not were manually assigned, using the designations of the Collins COBUILD (Collins Birmingham University International Language Database) dictionary (Sinclair, 1987). COBUILD marks each sense of each adjective with one of the labels QUALIT, CLASSIF, or COLOR, corresponding to gradable, non-gradable, and color adjectives. In cases where COBUILD supplies conflicting labels for different senses of a word, we either omitted that word or, if a sense were predominant, gave it the

label of that sense. In some cases, the word did not appear in COBUILD; these typically were descriptive compounds specific to the domain (e.g., *anti-takeover*, *over-the-counter*) and were in most cases marked as non-gradable adjectives. Overall, 453 of the 496 adjectives (91.33%) were assigned gradability labels by hand, while the remaining 53 words were discarded because they were misclassified as adjectives by the part-of-speech tagger (e.g., *such*) or because they could not be assigned a unique gradability label in accordance with COBUILD. Out of these words, 235 (51.88%) were manually classified as gradable adjectives, and 218 (48.12%) were classified as non-gradable adjectives.

Following the methodology of the preceding subsection, we recovered the inflection and modification indicators for these 453 adjectives, and trained both the unmodified and modified log-linear models repeatedly, using a randomly selected subset of 300 adjectives for training and 100 adjectives for testing. The entire cycle of selecting random test and training sets, fitting the model's coefficients, making predictions, and evaluating the predicted gradability labels is repeated 100 times, to ensure that the evaluation is not affected by a lucky (or unlucky) partition of the data between training and test sets. This procedure yields over the 453 adjectives gradability classifications with an average precision of 93.55% and average recall of 82.24% (in terms of the gradable words reported or recovered, respectively). The overall accuracy of the predicted gradability labels is 87.97%. These results were obtained with the modified log-linear model, which slightly outperformed the model that uses all three predictors (in that case, we obtained an average precision of 93.86%, average recall of 81.70%, and average overall accuracy of 87.70%). Figure 1 lists the gradability labels that were automatically assigned to one of the 100 random test sets using the modified prediction algorithm. We also assigned automatically labels to the entire set of 453 adjectives, using 4-fold cross-validation (repeatedly training on three-fourths of the 453 adjectives and testing on the rest). This resulted in precision of 94.15%, recall of 82.13%, and accuracy of 88.08% for the entire adjective set.

4 Subjectivity

The main motivation for the present paper is to examine the effect that information about an adjective's semantic orientation and gradability has on its probability of occurring in a subjective sentence (and hence on its quality as a subjectivity predictor). We first review related work on subjectivity recognition and then present our results.

4.1 Previous work on subjectivity recognition

In work by Wiebe, Bruce, and O'Hara (Wiebe et al., 1999; Bruce and Wiebe, 2000), a corpus of 1,001 sentences[3] of the Wall Street Journal TreeBank Corpus

---

[3]Compound sentences were manually segmented into their conjuncts, and each conjunct treated as a separate sentence.

(Marcus et al., 1993) was manually annotated with subjectivity classifications. Specifically, each sentence was assigned a subjective or objective classification, according to concensus tags derived by a statistical analysis of the classes assigned by three human judges (see (Wiebe et al., 1999) for further information). The total number of subjective sentences in the data is 486, and the total number of objective sentences is 515.

Bruce and Wiebe (2000) performed a statistical analysis of the assigned classifications, finding that adjectives are statistically significantly and positively correlated with subjective sentences in the corpus on the basis of the log-likelihood ratio test statistic G^2. The probability of a sentence being subjective, simply given that there is at least one adjective in the sentence, is 56%, even though there are more objective than subjective sentences in the corpus. In addition, Bruce and Wiebe identified a type of adjective that is indicative of subjective sentences: those Quirk et al. (1985) term *dynamic*, which "denote qualities that are thought to be subject to control by the possessor" (p. 434). Examples are "kind" and "careful". Bruce and Wiebe manually applied syntactic tests to identify dynamic adjectives in half of the corpus mentioned above. We include such adjectives in the analysis below, to assess whether additional lexical semantic features associated with subjectivity help improve predictability.

Wiebe et al. (1999) developed an automatic system to perform subjectivity tagging. In 10-fold cross validation experiments applied to the corpus described above, a probabilistic classifier obtained an average accuracy on subjectivity tagging of 72.17%, more than 20 percentage points higher than the baseline accuracy obtained by always choosing the more frequent class. A binary feature is included for each of the following: the presence in the sentence of a pronoun, an adjective, a cardinal number, a modal other than *will*, and an adverb other than *not*. They also included a binary feature representing whether or not the sentence begins a new paragraph. Finally, a feature was included representing co-occurrence of word tokens and punctuation marks with the subjective and objective classification. An analysis of the system showed that the adjective feature was important to realizing the improvements over the baseline accuracy. In this paper, we use the performance of the simple adjective feature as a baseline, and identify higher quality adjective features based on gradability and orientation.

4.2 Orientation and gradability as subjectivity predictors: Results

We measure the precision of a simple prediction method for subjectivity: a sentence is classified as subjective if at least one member of a set of adjectives S occurs in the sentence, and objective otherwise. By varying the set S (e.g., all adjectives, only gradable adjectives, only negatively oriented adjectives, etc.) we can assess the usefulness of the additional knowledge for predicting subjectivity.

For the present study, we use the set of all adjectives automatically identified in the corpus by Wiebe et al. (1999) (Section 4.1); the set of dynamic adjectives manually identified by Bruce and Wiebe (2000) (Section 4.1); the set of semantic orientation labels assigned by Hatzivassiloglou and McKeown (1997), both manually and automatically with our extension described in Section 2; and the set of gradability labels, both manually and automatically assigned according to the revised log-linear model of Section 3. We calculate results (shown in Table 3) for each of these sets of all adjectives, dynamic, oriented and gradable adjectives, as well as for unions and intersections of those sets. Note that these four sets have been extracted from comparable but different corpora (different years of the Wall Street Journal), therefore sometimes adjectives in one corpus may not be present in the other corpus, reducing the size of intersection sets. Also, for gradability, we worked with a sample set of 100 adjectives rather than all possible adjectives we could automatically calculate gradability values for, since our goal in the present work is to measure correlations between these sets and subjectivity, rather than building a system for predicting subjectivity for as many adjectives as possible.

In Table 3, the second column identifies S, the set of adjective types in question. The third column gives the number of subjective sentences that contain one or more instances of members of S, and the fourth column gives the same figure for objective sentences. Therefore these two columns together specify the coverage of the subjectivity indicator examined. The fifth column gives the conditional probability that a sentence is subjective, given that one or more instances of members of S appears. This is a precision metric that assesses feature quality: if instances of S appear, how likely is the sentence to be subjective? The last two columns contrast the observed conditional probability with the a priori probability of subjective sentences (i.e., chance; sixth column) and with the probability assigned by the baseline all-adjectives model (i.e., the first row in the table; seventh column).

The most striking aspect of these results is that all sets involving dynamic adjectives, positive or negative polarity, or gradability are better predictors of subjective sentences than the class of adjectives as a whole. Five of the sets are at least 25 points better (L14, L16, L21, L23, and L24); four others are at least 20 points better (L2, L9, L13, and L15); and five others are at least 15 points better (L4, L11, L18, L20, and L22). In most of these cases, the difference between these predictors and all adjectives is statistically significant[4] at the 5% level or less; almost all of these predictors offer statistically significantly better than even odds in predicting subjectivity correctly. In many cases where statistical significance

---

[4]We applied a chi-square test on the 2×2 cross-classification table (Fleiss, 1981).

	Adjective Set $\mathcal{S}$	# Subj Sents with $(s \in S)^+$	# Obj Sents with $(s \in S)^+$	P(Subj Sent | $(s \in S)^+$)	Significance Against majority	Significance Against all adjs
L1.	All Adjectives	403	321	0.56	0.0041	N/A
L2.	Dynamic Adjectives	92	32	0.74	$1.1989 \cdot 10^{-7}$	$1.6369 \cdot 10^{-4}$
L3.	Pol+, man	138	87	0.61	0.0007	0.1546
L4.	Pol−, man	79	37	0.67	0.0001	0.0158
L5.	Pol+ ∪ Pol−, man	197	114	0.63	$6.9191 \cdot 10^{-6}$	0.0260
L6.	Grad, man	193	115	0.63	$1.9633 \cdot 10^{-5}$	0.0440
L7.	Not Grad, man	172	147	0.54	0.1084	0.6496
L8.	Pol+, auto	121	79	0.60	0.0026	0.2537
L9.	Pol−, auto	61	21	0.74	$1.1635 \cdot 10^{-5}$	0.0017
L10.	Pol+ ∪ Pol−, auto	170	95	0.64	$8.5888 \cdot 10^{-6}$	0.0202
L11.	Grad, auto	30	14	0.68	0.0166	0.1418
L12.	Not Grad, auto	63	51	0.55	0.2079	0.9363
L13.	Dyn Adjs ∩ $\mathcal{S}$ of L5.	51	19	0.73	0.0001	0.0081
L14.	Dyn Adjs ∩ $\mathcal{S}$ of L6.	39	8	0.83	$8.0397 \cdot 10^{-6}$	$4.3671 \cdot 10^{-4}$
L15.	Dyn Adjs ∩ $\mathcal{S}$ of L10.	50	19	0.72	0.0002	0.0103
L16.	Dyn Adjs ∩ $\mathcal{S}$ of L11	7	2	0.78	0.1582	0.3220
L17.	Grad ∩ Pol+, man	90	58	0.61	0.0070	0.2891
L18.	Grad ∩ Pol−, man	35	16	0.69	0.0080	0.09711
L19.	Grad ∩ (Pol+ ∪ Pol−), man	119	71	0.63	0.0005	0.1000
L20.	Grad ∩ Pol+, auto	13	6	0.68	0.1376	0.3833
L21.	Grad ∩ Pol−, auto	2	0	1.00	0.4556	0.5838
L22.	Grad ∩ (Pol+ ∪ Pol−), auto	15	6	0.71	0.0636	0.2255
L23.	Dyn Adjs ∩ $\mathcal{S}$ of L22.	4	0	1.00	0.1203	0.2019
L24.	Dyn Adjs ∩ $\mathcal{S}$ of L19.	24	5	0.83	0.0006	0.0070

Key: $(s \in S)^+$: one or more instances of members of $\mathcal{S}$. *Pol+*: positive polarity. *Pol−*: negative polarity.
Grad: gradable. *Dyn*: dynamic. *Man*: manually identified. *Auto*: automatically identified.

Table 3: Subjectivity prediction results.

could not established this is due to small counts, caused by the small size of the set of adjectives automatically labeled for gradability.

It is also important to note that, in most cases, the automatically-classified adjectives are comparable or better predictors of subjective sentences than the manually-assigned ones. Comparing the automatically generated classes with the manually identified ones, the positive polarity set decreases by 1 percentage point (L3 and L8), while the negative polarity set increases by 7 points (L4 and L9), and the gradable set increases by 5 percentage points (L6 and L11). Among the intersection sets, in two cases the results are lower for the computer-generated sets (L13/L15 and L14/L16), but in the other 4 cases, the results are higher (L17/L20, L18/L21, L19/L2, L24/L23).

Finally, the table shows that, in most cases, predictability improves or at worst remains essentially the same as additional lexical features are considered. For the set of dynamic adjectives, the predictability is 74% (L2), and improves in 4 of the 6 cases in which it is intersected with other sets (L14, L16, L23, and L24). For the other two (L13 and L15), predictability is only 1 or 2 points lower (not statistically significant). For the manually assigned polarity and gradability sets, in one case predictability is lower (L17 < L6), but in the other cases

it remains the same or improves. The results are even better for the automatically assigned polarity and gradability sets: predictability improves when both features are considered in all but one case, when predictability remains the same (L20 > L8; L21 > L9; L22 > L10; and L11 ≤ L20, L21, and L22).

5 Conclusion and Future Work

This paper presents an analysis of different adjective features for predicting subjectivity, showing that they are more precise than those previously used for this task. We establish that lexical semantic features such as semantic orientation and gradability determine in large part the subjectivity status of sentences in which they appear. We also present an automatic method for extracting gradability values reliably, complementing earlier work on semantic orientation and dynamic adjectives.

In addition to finding more precise features for automatic subjectivity recognition, this kind of analysis could help efforts to encode subjective features in ontologies such as those described in (Knight and Luk, 1994; Mahesh and Nirenburg, 1995; Hovy, 1998). These ontologies are useful for many NLP tasks, such as machine translation, word-sense disambiguation, and generation. Some subjective features are included in existing ontologies (for example, Mikrokosmos (Mahesh and

Nirenburg, 1995) includes attitude slots). Our corpus-based methods could help in identifying more or extending their coverage.

To be able to use automatic subjectivity recognition in text-processing applications, good clues of subjectivity must be found. The features developed in this paper are not only good clues of subjectivity, they can be identified automatically from corpora (see (Hatzivassiloglou and McKeown, 1997), and Section 3 in the present paper). In fact, the results in Table 3 show that the predictability of the automatically determined gradability and polarity sets is better than or at least comparable to the predictability of the manually determined sets. Thus, the oriented and gradable adjectives in the particular application genre can be identified for use in subjectivity recognition.

Our efforts in this paper are largely exploratory, aiming to establish correlations among the various features examined. In related work, we have begun to incorporate the features developed here into systems for recognizing flames and mining reviews in Internet forums, extending subjectivity judgments from the sentence to the document level. In addition, we are seeking ways to extend the orientation and gradability methods so that individual word occurrences, rather than word types, are characterized as oriented or gradable. We also plan to incorporate the new features presented here in machine learning models for the prediction of subjectivity (e.g., (Wiebe et al., 1999)) and test their interactions with other proposed features.

Acknowledgments

References

Douglas M. Bates and Donald G. Watts. 1988. *Nonlinear Regression Analysis and its Applications*. Wiley, New York.

Edwin L. Battistella. 1990. *Markedness: The Evaluative Superstructure of Language*. State University of New York Press, Albany, New York.

Rebecca Bruce and Janyce Wiebe. 2000. Recognizing subjectivity: A case study of manual tagging. *Natural Language Engineering*, **6**(2).

Kenneth W. Church. 1988. A stochastic parts program and noun phrase parser for unrestricted text. In *Proceedings of the Second Conference on Applied Natural Language Processing (ANLP-88)*, pages 136–143, Austin, Texas, February. Association for Computational Linguistics.

Joseph L. Fleiss. 1981. *Statistical Methods for Rates and Proportions*. Wiley, New York, 2nd edition.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL*, pages 174–181, Madrid, Spain, July. Association for Computational Linguistics.

Eduard Hovy. 1998. Combining and standardizing large-scale practical ontologies for machine translation and other uses. In *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC)*, Granada, Spain.

Jerrold J. Katz. 1972. *Semantic Theory*. Harper and Row, New York.

Kevin Knight and Steve K. Luk. 1994. Building a large-scale knowledge base for machine translation. In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94)*, volume 1, pages 773–778, Seattle, Washington, July–August. American Association for Artificial Intelligence.

Adrienne Lehrer. 1974. *Semantic Fields and Lexical Structure*. North Holland, Amsterdam and New York.

John Lyons. 1977. *Semantics*, volume 1. Cambridge University Press, Cambridge, England.

K. Mahesh and S. Nirenburg. 1995. A situated ontology for practical NLP. In *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, 14th International Joint Conference on Artificial Intelligence (IJCAI-95)*, Montréal, Canada, August.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, **19**(2):313–330, June.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman, London and New York.

Thomas J. Santner and Diane E. Duffy. 1989. *The Statistical Analysis of Discrete Data*. Springer-Verlag, New York.

Edward Sapir. 1944. On grading: A study in semantics. *Philosophy of Science*, **2**:93–116. Reprinted in (Sapir, 1949).

Edward Sapir. 1949. *Selected Writings in Language, Culture and Personality*. University of California Press, Berkeley, California. Edited by David G. Mandelbaum.

John M. Sinclair (editor in chief). 1987. *Collins COBUILD English Language Dictionary*. Collins, London.

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