More than Words:
Syntactic Packaging and Implicit Sentiment

Stephan Greene[*]
ATG, Inc.
1111 19th St, NW Suite 600
Washington, DC 20036
sgreene@atg.com

Philip Resnik
Linguistics / UMIACS CLIP Laboratory
University of Maryland
College Park, MD 20742
resnik@umiacs.umd.edu

Abstract

Work on sentiment analysis often focuses on the words and phrases that people use in overtly opinionated text. In this paper, we introduce a new approach to the problem that focuses not on lexical indicators, but on the syntactic "packaging" of ideas, which is well suited to investigating the identification of *implicit sentiment*, or perspective. We establish a strong predictive connection between linguistically well motivated features and implicit sentiment, and then show how computational approximations of these features can be used to improve on existing state-of-the-art sentiment classification results.

1 Introduction

As Pang and Lee (2008) observe, the last several years have seen a "land rush" in research on sentiment analysis and opinion mining, with a frequent emphasis on the identification of opinions in evaluative text such as movie or product reviews. However, sentiment also may be carried implicitly by statements that are not only non-evaluative, but not even visibly subjective. Consider, for example, the following two descriptions of the same (invented) event:

1(a) On November 25, a soldier veered his jeep into a crowded market and killed three civilians.

 (b) On November 25, a soldier's jeep veered into a crowded market, causing three civilian deaths.

Both descriptions appear on the surface to be objective statements, and they use nearly the same words. Lexically, the sentences' first clauses differ only in the difference between *'s* and *his* to express the relationship between the soldier and the jeep, and in the second clauses both *kill* and *death* are terms with negative connotations, at least according to the General Inquirer lexicon (Stone, 1966). Yet the descriptions clearly differ in the feelings they evoke: if the soldier were being tried for his role in what happened on November 25, surely the prosecutor would be more likely to say (1a) to the jury, and the defense attorney (1b), rather than the reverse.[1]

Why, then, should a description like (1a) be perceived as less sympathetic to the soldier than (1b)? If the difference is not in the words, it must be in the way they are put together; that is, the *structure* of the sentence. In Section 2, we offer a specific hypothesis about the connection between structure and implicit sentiment: we suggest that the relationship is mediated by a set of "grammatically relevant" semantic properties well known to be important cross-linguistically in characterizing the interface between syntax and lexical semantics. In Section 3, we validate this hypothesis by means of a human ratings study, showing that these properties are highly predictive of human sentiment ratings. In Section 4, we introduce *observable proxies for underlying semantics* (OPUS), a practical way to approximate the relevant semantic properties automatically as features in a supervised learning setting. In Section 5, we show that these features improve on the existing state of the art in automatic sentiment classification. Sec-

---

[*]This work was done while the first author was a student in the Department of Linguistics, University of Maryland.

[1]We refer readers not sharing this intuition to Section 3.

tions 6 and 7 discuss related work and summarize.

2 Linguistic Motivation

Verbal descriptions of an event often carry along with them an underlying attitude toward what is being described. By framing the same event in different ways, speakers or authors "select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation" (Entman, 1993, p. 52). Clearly lexical choices can accomplish this kind of selection, e.g. choosing to describe a person as a *terrorist* rather than a *freedom fighter*, or referencing *killer whales* rather than *orcas*.[2] Syntactic choices can also have framing effects. For example, Ronald Reagan's famous use of the passive construction, "Mistakes were made" (in the context of the Iran-Contra scandal), is a classic example of framing or spin: used without a *by*-phrase, the passive avoids identifying a causal agent and therefore sidesteps the issue of responsibility (Broder, 2007). A toddler who says "My toy broke" instead of "I broke my toy" is employing the same linguistic strategy.

Linguists have long studied syntactic variation in descriptions of the same event, often under the general heading of syntactic diathesis alternations (Levin, 1993; Levin and Hovav, 2005). This line of research has established a set of semantic properties that are widely viewed as "grammatically relevant" in the sense that they enable generalizations about syntactic "packaging" of meaning within (and across) the world's languages. For example, the verb *break* in English participates in the causative-inchoative alternation (causative event *X broke Y* can also be expressed without overt causation as *Y broke*), but the verb *climb* does not (*X* also causes the event in *X climbed Y*, but that event cannot be expressed as *Y climbed*). These facts about participation in the alternation turn out to be connected with the fact that a breaking event entails a *change of state* in *Y* but a climbing event does not. Grammatically relevant semantic properties of events and their

participants — causation, change of state, and others — are central not only in theoretical work on lexical semantics, but in computational approaches to the lexicon, as well (e.g. (Pustejovsky, 1991; Dorr, 1993; Wu and Palmer, 1994; Dang et al., 1998)).

The approach we propose draws on two influential discussions about grammatically relevant semantic properties in theoretical work on lexical semantics. First, Dowty (1991) characterizes grammatically relevant properties of a verb's arguments (e.g. subject and object) via inferences that follow from the meaning of the verb. For example, expressions like *X murders Y* or *X interrogates Y* entail that subject X caused the event.[3] Second, Hopper and Thompson (1980) characterize "semantic transitivity" using similar properties, connecting semantic features to morphosyntactic behavior across a wide variety of languages.

Bringing together Dowty with Hopper and Thompson, we find 13 semantic properties organized into three groups, corresponding to the three components of a canonical transitive clause, expressed as *X verb Y* in English.[4] *Properties associated with X* involve volitional involvement in the event or state, causation of the event, sentience/awareness and/or perception, causing a change of state in *Y*, kinesis or movement, and existence independent of the event. *Properties associated with the event or state conveyed by the verb* include aspectual features of telicity (a defined endpoint) and punctuality (the latter of which may be inversely related to a property known as incremental theme). *Properties associated with Y* include affectedness, change of state, (lack of) kinesis or movement, and (lack of) existence independent of the event.

Now, observe that this set of semantic properties involves many of the questions that would naturally help to shape one's opinion about the event described by *veer* in (1). Was anyone or anything affected by what took place, and to what degree? Did the event just happen or was it caused? Did the event reach a defined endpoint? Did participation in

---

[2]Supporters of an endangered species listing in Puget Sound generally referred to the animals as *orcas*, while opponents generally said *killer whales* (Harden, 2006).

[3]Kako (2006) has verified that people make these inferences based on X's syntactic position even when a semantically empty nonsense verb is used.

[4]We are deliberately sidestepping the choice of terminology for X and Y, e.g. proto-Patient, theme, etc.

the event involve conscious thought or intent? Our hypothesis is that the syntactic aspects of "framing", as characterized by Entman, involve manipulation of these semantic properties, even when overt opinions are not being expressed. That is, we propose a connection between syntactic choices and implicit sentiment mediated by the very same semantic properties that linguists have already identified as central when connecting surface expression to underlying meaning more generally.

3 Empirical Validation

We validated the hypothesized connection between implicit sentiment and grammatically relevant semantic properties using psycholinguistic methods, by varying the syntactic form of event descriptions, and showing that the semantic properties of descriptions do indeed predict perceived sentiment.[5]

3.1 Semantic property ratings

Materials. Stimuli were constructed using 11 verbs of killing, which are widely viewed as prototypical for the semantic properties of interest here (Lemmens, 1998): *X killed Y* normally involves conscious, intentional causation by *X* of a kinetic event that causes a (rather decisive and clearly terminated!) change of state in *Y*. The verbs comprise two classes: the "transitive" class, involving externally caused change-of-state verbs (*kill*, *slaughter*, *assassinate*, *shoot*, *poison*), and the "ergative" class (*strangle*, *smother*, *choke*, *drown*, *suffocate*, *starve*), within which verbs are internally caused (McKoon and MacFarland, 2000) or otherwise emphasize properties of the object. Variation of syntactic description involved two forms: a transitive syntactic frame with a human agent as subject ("transitive form", 2a), and a nominalization of the verb as subject and the verb *kill* as the predicate ("nominalized form", 2b).

2(a) The gunmen shot the opposition leader
 (b) The shooting killed the opposition leader

Participants and procedure. A set of 18 volunteer participants, all native speakers of English, were presented with event descriptions and asked to answer questions probing both Dowty's proto-role

properties as well as Hopper and Thompson's semantic transitivity components, responding via ratings on a 1-to-7 scale. For example, the questions probing volition were: "*In this event*, how likely is it that ⟨subject⟩ chose to be involved?", where ⟨subject⟩ was *the gunmen* and *the shooting*, for 2(a-b), respectively.[6]

3.2 Sentiment ratings

Materials. We used the materials above to construct short, newspaper-like paragraphs, each one accompanied by a "headline" version of the same syntactic descriptions used above. For example, given this paragraph:

> A man has been charged for the suffocation of a woman early Tuesday morning. City police say the man suffocated the 24-year-old woman using a plastic garbage bag. The woman, who police say had a previous relationship with her attacker, was on her way to work when the incident happened. Based on information provided by neighbors, police were able to identify the suspect, who was arrested at gunpoint later the same day.

the three alternative headlines would be:

3(a) Man suffocates 24-year old woman
 (b) Suffocation kills 24-year-old woman
 (c) 24-year-old woman is suffocated

Some paragraphs were based on actual news stories.[7] In all paragraphs, there is an obvious nominal referent for both the perpetrator and the victim, it is clear that the victim dies, and the perpetrator in the scenario is responsible for the resulting death directly rather than indirectly (e.g. through negli-

---

[5]Full details and materials in Greene (2007).

[6]Standard experimental design methods were followed with respect to counterbalancing, block design, and distractor stimuli; for example, no participant saw more than one of 2(a) or 2(b), and all participants saw equal numbers of transitive and nominalized descriptions. The phrase *In this event* was repeated in each question and emphasized visually in order to encourage participants to focus on the particular event described in the sentence, rather than on the entities or events denoted in general.

[7]In those cases no proper names were used, to avoid any inadvertent emotional reactions or legal issues, although the descriptions retained emotional impact because we wanted readers to have some emotional basis with which to judge the headlines.

gence).[8] The stem of the nominalization always appeared in the event description in either verbal or nominal form.

Participants and procedure. A set of 31 volunteers, all native speakers of English, were presented with the paragraph-length descriptions and accompanying headlines. As a measure of sentiment, participants were asked to rate headlines on a 1-to-7 scale with respect to how sympathetic they perceive the headline to be toward the perpetrator. For example, given the paragraph and one of the associated headlines in (3), a participant would be asked to rate "How sympathetic or unsympathetic is *this headline* to the man?"[9]

3.3 Analysis and discussion

Unsurprisingly, but reassuringly, an analysis of the sentiment ratings yields a significant effect of syntactic form on sympathy toward the perpetrator ($F(2, 369) = 33.902, p < .001$), using a mixed model ANOVA run with the headline form as fixed effect. The transitive form of the headline yielded significantly lower sympathy ratings than the nominalized or passive forms in pairwise comparisons (both $p < .001$). We have thus confirmed empirically that Reagan's "Mistakes were made" was a wise choice of phrasing on his part.

More important, we are now in a position to examine the relationship between syntactic forms and perceived sentiment in more detail. We performed regression analyses treating the 13 semantic property ratings plus the identity of the verb as independent variables to predict sympathy rating as a dependent variable, using the 24 stimulus sentences that bridged both collections of ratings.[10] Considering semantic properties individually, we find that volition has the strongest correlation with sympathy (a negative correlation, with $r = -.776$), followed by sentience ($r = -.764$) and kinesis/movement ($r = -.751$). Although performing a multiple regression with all variables for this size dataset is impossible, owing to overfitting (as a rule of thumb, 5 to 10 observed items are necessary per each independent variable), a multiple regression involving verb, volition, and telicity as independent variables yields $R = .88$, $R^2 = .78$ ($p < .001$). The value for adjusted R^2, which explicitly takes into account the small number of observations, is 74.1.

In summary, then, this ratings study confirms the influence of syntactic choices on perceptions of implicit sentiment. Furthermore, it provides support for the idea that this influence is mediated by "grammatically relevant" semantic properties, demonstrating that these accounted for approximately 75% of the variance in implicit sentiment expressed by alternative headlines describing the same event.

4 Observable Approximation

Thus far, we have established a predictive connection between syntactic choices and underlying or implicit sentiment, mediated by grammatically relevant semantic properties. In an ideal world, we could harness the predictive power of those properties by using volition, causation, telicity, etc. as features for regression or classification in sentiment prediction tasks. Unfortunately, the properties are not directly observable, and neither automatic annotators nor labeled training data currently exist.

We therefore pursue a different strategy, which we refer to as *observable proxies for underlying semantics* (OPUS). It can be viewed as a middle ground between relying on construction-level syntactic distinctions (such as the 3-way transitive, nominalized subject, passive distinction in Section 3) and annotation of fine-grained semantic properties. The key idea is to use observable grammatical relations, drawn from the usages of terms determined to be relevant to a domain, as proxies for the underlying semantic properties that gave rise to their syntactic realization using those relations. Automatically cre-

---

[8]An alert reader may observe that headlines with nominalized subjects using the verb *kill* require some other nominalization, so they don't say "Killing kills victim". For these cases in the data, an appropriate nominalization drawn from the event description was used (e.g., *explosion*).

[9]Again, standard experimental design methods were used with respect to block design, distractor stimuli, etc. The phrase *this headline* was emphasized to stress that it is the headline being rated, not the story. A second question rating sympathy toward the victim was also asked in each case, as an additional distractor.

[10]These involved only the transitive and nominalized forms, because many of the questions were inapplicable to the passive form. Since the two ratings studies involved different subject pools, regression models were run over the mean values of each observation in the experimental data.

ated features based on those observable proxies are then used in classification as described in Section 5.

In order to identify the set T of terms relevant to a particular document collection, we adopt the relative frequency ratio (Damerau, 1993), $R(t) = R_{\text{domain}}^t / R_{\text{reference}}^t$, where $R_c^t = \frac{f_c^t}{N_c}$ is the ratio of term t's frequency in corpus c to the size N_c of that corpus. $R(t)$ is a simple but effective comparison of a term's prevalence in a particular collection as compared to a general reference corpus. We used the British National Corpus as the reference because it is both very large and representative of text from a wide variety of domains and genres. The threshold of $R(t)$ permitting membership in T is an experimental parameter.

OPUS features are defined in terms of syntactic dependency relations involving terms in T. Given a set D of syntactic dependency relations, features are of the form $t : d$ or $d : t$, with $d \in D, t \in T$. That is, they are term-dependency pairs extracted from term-dependency-term dependency tuples, preserving whether the term is the head or the dependent in the dependency relation. In addition, we add two construction-specific features: TRANS:v, which represents verb v in a canonical, syntactically transitive usage, and NOOBJ:v, present when verb v is used without a direct object.[11]

Example 4 shows source text (bolded clause in 4a), an illustrative subset of parser dependencies (4b), and corresponding OPUS features (4c):

4(a) Life Without Parole does not eliminate the risk that **the prisoner will murder a guard, a visitor, or another inmate**.
 (b) nsubj(murder, prisoner); aux(murder, will); dobj(murder, guard)
 (c) TRANS:murder, murder:nsubj, nsubj:prisoner, murder:aux, aux:will, murder:dobj, dobj:guard

Intuitively the presence of TRANS:murder suggests the entire complex of semantic properties discussed in Section 2, bringing together the impliciation of volition, causation, etc. on the part of *prisoner* (as does nsubj:prisoner), affectedness and change of state on the part of *guard* (as does dobj:guard), and so forth.

The NOOBJ features can capture a habitual reading, or in some cases a detransitivizing effect associated with omission of the direct object (Olsen and Resnik, 1997). The bold text in (5) yields NOOBJ:kill as a feature.

5(a) At the same time, we should never ignore the risks of allowing the inmate **to kill again**.

In this case, omitting the direct object decreases the extent to which the killing event is interpreted as telic, and it eliminates the possibility of attributing change-of-state to a specific affected object (much like "Mistakes were made" avoids attributing cause to a specified subject), placing the phrasing at a less "semantically transitive" point on the transitivity continuum (Hopper and Thompson, 1980). Some informants find a perceptible increase in negative sentiment toward *inmate* when the sentence is phrased as in 5(b):

5(b) At the same time, we should never ignore the risks of allowing the inmate **to kill someone again**.

5 Computational Application

Having discussed linguistic motivation, empirical validation, and practical approximation of semantically relevant features, we now present two studies demonstrating their value in sentiment classification. For the first study, we have constructed a new data set particularly well suited for testing our approach, based on writing about the death penalty. In our second study, we make a direct comparison with prior state-of-the-art classification using the Bitter Lemons corpus of Lin et al. (2006).

5.1 Predicting Opinions of the Death Penalty

Corpus. We constructed a new corpus for experimentation on implicit sentiment by downloading the contents of pro- and anti-death-penalty Web sites and manually checking, for a large subset, that the viewpoints expressed in documents were as expected. The collection, which we will refer to as the DP corpus, comprises documents from five pro-death-penalty sites and three anti-death-penalty sites, and the corpus was engineered to have an even balance, 596 documents per side.[12]

---

[11] We parsed English text using the Stanford parser.

[12] Details in Greene (2007).

Frequent bigram baseline. We adopted a supervised classification approach based on word n-gram features, using SVM classification in the WEKA machine learning package. In initial exploration using both unigrams and bigrams, and using both word forms and stems, we found that performance did not differ significantly, and chose stemmed bigrams for our baseline comparisons. In order to control for the difference in the number of features available to the classifier in our comparisons, we use the N most frequent stemmed bigrams as the baseline feature set where N is matched to number of OPUS features used in the comparison condition.

OPUS-kill verbs: OPUS features for manually selected verbs. We created OPUS features for 14 verbs — those used in Section 3, plus *murder*, *execute*, and *stab* and their nominalizations (including both event and *-er* nominals, e.g. both *killing* and *killer*) — generating $N = 1016$ distinct features.

OPUS-domain: OPUS features for domain-relevant verbs. We created OPUS features for the 117 verbs for which the relative frequency ratio was greater than 1. This list includes many of the *kill* verbs we used in Section 3, and introduces, among others, many transitive verbs describing acts of physical force (e.g. *rape*, *rob*, *steal*, *beat*, *strike*, *force*, *fight*) as well as domain-relevant verbs such as *testify*, *convict*, and *sentence*. Included verbs near the borderline included, for example, *hold*, *watch*, *allow*, and *try*. Extracting OPUS features for these verbs yielded $N = 7552$ features.

Evaluation. Cross-validation at the document level does not test what we are interested in, since a classifier might well learn to bucket documents according to Web site, not according to pro- or anti-death-penalty sentiment. To avoid this difficulty, we performed site-wise cross-validation. We restricted our attention to the two sites from each perspective with the most documents, which we refer to as *pro1*, *pro2*, *anti1*, and *anti2*, yielding 4-fold cross-validation. Each fold $f_{train,test}$ is defined as containing all documents from one pro and one anti site for training, using all documents from the remaining pro and anti sites for testing. So, for example, fold $f_{11,22}$ uses all documents from *pro1* and *anti1* in training, and all documents from *pro2* and

Condition	N features	SVM accuracy
Baseline	1016	68.37
OPUS-kill verbs	1016	82.09
Baseline	7552	71.96
OPUS-domain	7552	88.10

Table 1: Results for 4-fold site-wise cross-validation using the DP corpus

Condition	N features	SVM accuracy
Baseline	1518	55.95
OPUS-frequent verbs	1518	55.95
OPUS-kill verbs	1062	66.67

Table 2: DP corpus comparison for OPUS features based on frequent vs. domain-relevant verbs

anti2 for testing.[13] As Table 1 shows, OPUS features provide substantial and statistically significant gains ($p < .001$).

As a reality check to verify that it is domain-relevant verb usages and the encoding of events they embody that truly drives improved classification, we extracted OPUS features for the 14 most frequent verbs found in the DP Corpus that were *not* in our manually created list of kill verbs, along with their nominalizations. Table 2 shows the results of a classification experiment using a single train-test split, training on 1062 documents from *pro1*, *pro2*, *anti1*, *anti2* and testing on 84 test documents from the significantly smaller remaining sites. Using OPUS features for the most frequent non-*kill* verbs fails to beat the baseline, establishing that it is not simply term frequency, the presence of particular grammatical relations, or a larger feature set that the *kill*-verb OPUS model was able to exploit, but rather the properties of event encodings involving the *kill* verbs themselves.

5.2 Predicting Points of View in the Israeli-Palestinian Conflict

In order to make a direct comparison here with prior state-of-the-art work on sentiment analysis, we report on sentiment classification using OPUS features in experiments using a publicly available corpus involving opposing perspectives, the Bitter Lemons

---

[13]Site (# of documents): pro1= clarkprosecutor.org (437), pro2= prodeathpenalty.com (117), anti1= deathpenaltyinfo.org (319), anti2= nodeathpenalty.org (212)

(hence BL) corpus introduced by Lin et al. (2006).

Corpus. The Bitter Lemons corpus comprises essays posted at `www.bitterlemons.org`, which, in the words of the site, "present Israeli and Palestinian viewpoints on prominent issues of concern". As a corpus, it has a number of interesting properties. First, its topic area is one of significant interest and considerable controversy, yet the general tenor of the web site is one that eschews an overly shrill or extreme style of writing. Second, the site is organized in terms of issue-focused weekly editions that include essays with contrasting viewpoints from the site's two editors, plus two essays, also contrasting, from guest editors. This creates a natural balance between the two sides and across the subtopics being discussed. The BL corpus as prepared by Lin et al. contains 297 documents from each of the Israeli and Palestinian viewpoints, averaging 700-800 words in length.

Lin et al. classifiers. Lin et al. report results on distinguishing Israeli vs. Palestinian perspectives using an SVM classifier, a naive Bayes classifier NB-M using maximum a posteriori estimation, and a naive Bayes classifier NB-B using full Bayesian inference. (Document perspectives are labeled clearly on the site.) We continue to use the WEKA SVM classifier, but compare our results to both their SVM and NB-B, since the latter achieved their best results.

OPUS features. As in Section 5.1, we experimented with OPUS features driven by automatically extracted lists of domain-relevant verbs. For these experiments, we included domain-relevant nouns, and we varied a threshold ρ for the relative frequency ratio, including only terms for which $\log(R(t)) > \rho$. In addition, we introduced a general filter on OPUS features, eliminating syntactic dependency types that do not usefully reflect semantically relevant properties: det, predet, preconj, prt, aux, auxpas, cc, punct, complm, mark, rel, ref, expl.

Evaluation. Lin et al. describe two test scenarios. In the first, referred to as Test Scenario 1, they trained on documents written by the site's guests, and tested on documents from the site's editors. Test Scenario 2 represents the reverse, training on documents from the site editors and testing on documents
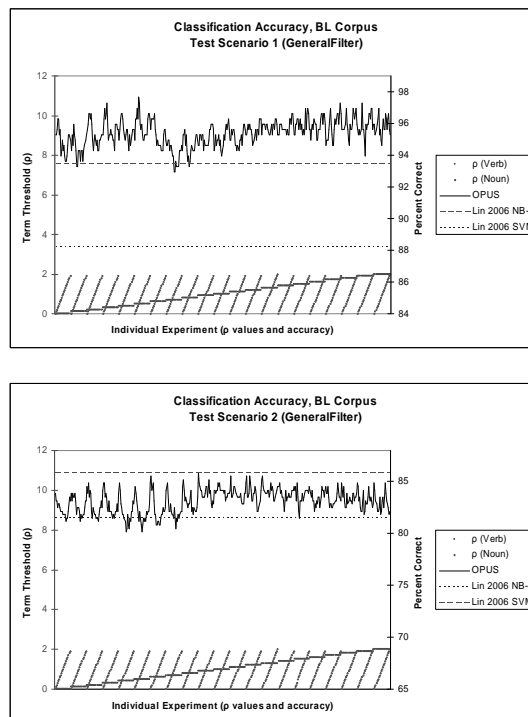

Figure 1: Results on the Bitter Lemons corpus

from guest authors. As in our site-wise cross validation for the DP corpus, this strategy ensures that what is being tested is classification according to the viewpoint, not author or topic.

Figure 1 (top) summarizes a large set of experiments for Test Scenario 1, in which we varied the values of ρ for verbs and nouns. Each experiment, using a particular $\langle \rho(\text{verbs}), \rho(\text{nouns}) \rangle$, corresponds to a vertical strip on the x-axis. The points on that strip include the ρ values for verbs and nouns, measured by the scale on the y-axis at the left of the figure; the accuracy of Lin et al.'s SVM (88.22% accuracy, constant across all our variations); the accuracy of Lin et al.'s NB-B classifier (93.46% accuracy, constant across all our variations), and the accuracy of our SVM classifier using OPUS features, which varies depending on the ρ values. Across 423 experiments, our average accuracy is 95.41%, with the best accuracy achieved being 97.64%. Our classifier underperformed NB-B slightly, with accuracies from 92.93% to 93.27%, in just 8 of the 423 experiments.

Figure 1 (bottom) provides a similar summary for

experiments in Test Scenario 2. The first thing to notice is that accuracy for all methods is lower than for Test Scenario 1. This is not terribly surprising: it is likely that training a classifier on the more uniform authorship of the editor documents builds a model that generalizes less well to the more diverse authorship of the guest documents (though accuracy is still quite high). In addition, the editor-authored documents comprise a smaller training set, consisting of 7,899 sentences, while the guest documents have a total of 11,033 sentences, a 28% difference. In scenario 2, we obtain average accuracy across experiments of 83.12%, with a maximum of 85.86%, in this case outperforming the 81.48% obtained by Lin's SVM fairly consistently, and in some cases approaching or matching NB-B at 85.85%.

6 Related Work

Pang and Lee's (2008) excellent monograph provides a thorough, well organized, and relatively recent description of computational work on sentiment, opinion, and subjectivity analysis.

The problem of classifying underlying sentiment in statements that are not overtly subjective is less studied within the NLP literature, but it has received some attention in other fields. These include, for example, research on content analysis in journalism, media studies, and political economy (Gentzkow and Shapiro, 2006a; Gentzkow and Shapiro, 2006b; Groseclose and Milyo, 2005; Fader et al., 2007); automatic identification of customer attitudes for business e-mail routing (Durbin et al., 2003). And, of course, the study of perceptions in politics and media bears a strong family resemblance to real-world marketing problems involving reputation management and business intelligence (Glance et al., 2005).

Within computational linguistics, what we call implicit sentiment was introduced as a topic of study by Lin et al. (2006) under the rubric of identifying perspective, though similar work had begun earlier in the realm of political science (e.g. (Laver et al., 2003)). Other recent work focusing on the notion of perspective or ideology has been reported by Martin and Vanberg (2008) and Mullen and Malouf (2008).

Among prior authors, Gamon's (2004) research is perhaps closest to the work described here, in that he uses some features based on a sentence's logical form, generated using a proprietary system. However, his features are templatic in nature in that they do not couple specific lexical entries with their logical form. Hearst (1992) and Mulder et al. (2004) describe systems that make use of argument structure features coupled with lexical information, though neither provides implementation details or experimental results.

In terms of computational experimentation, work by Thomas et al. (2006), predicting yes and no votes in corpus of United States Congressional floor debate speeches, is quite relevant. They combined SVM classification with a min-cut model on graphs in order to exploit both direct textual evidence and constraints suggested by the structure of Congressional debates, e.g. the fact that the same individual rarely gives one speech in favor of a bill and another opposing it. We have extend their method to use OPUS features in the SVM and obtained significant improvements over their classification accuracy (Greene, 2007; Greene and Resnik, in preparation).

7 Conclusions

In this paper we have introduced an approach to implicit sentiment motivated by theoretical work in lexical semantics, presenting evidence for the role of semantic properties in human sentiment judgments. This research is, to our knowledge, the first to draw an explicit and empirically supported connection between theoretically motivated work in lexical semantics and readers' perception of sentiment. In addition, we have reported positive sentiment classification results within a standard supervised learning setting, employing a practical first approximation to those semantic properties, including positive results in a direct comparison with the previous state of the art.

Because we computed OPUS features for opinionated as well as non-evaluative language in our corpora, obtaining overall positive results, we believe these features may also improve conventional opinion labeling for subjective text. This will be investigated in future work.

Acknowledgments

References

John Broder. 2007. Familiar fallback for officials: 'mistakes were made'. *New York Times*. March 14.

F. J. Damerau. 1993. Generating and evaluating domain-oriented multi-word terms from texts. *Information Processing and Management*, 29:433–447.

Hoa Trang Dang, Karin Kipper, Martha Palmer, and Joseph Rosenzweig. 1998. Investigating Regular Sense Extensions Based on Intersective Levin Classes. In *ACL/COLING 98*, pages 293–299, Montreal, Canada, August 10–14.

Bonnie J. Dorr. 1993. *Machine Translation: A View from the Lexicon*. The MIT Press, Cambridge, MA.

David Dowty. 1991. Thematic Proto-Roles and Argument Selection. *Language*, 67:547–619.

S. D. Durbin, J. N. Richter, and D. Warner. 2003. A system for affective rating of texts. In *Proc. 3rd Workshop on Operational Text Classification, KDD-2003*.

Robert M. Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58.

Anthony Fader, Dragomir R. Radev, Michael H. Crespin, Burt L. Monroe, Kevin M. Quinn, and Michael Colaresi. 2007. MavenRank: Identifying influential members of the US Senate using lexical centrality. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Michael Gamon. 2004. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In *Proc. COLING*.

M. Gentzkow and J. Shapiro. 2006a. Media bias and reputation. *Journal of Political Economy*, 114:280–316.

M. Gentzkow and J. Shapiro. 2006b. What drives media slant? Evidence from U.S. newspapers. http://ssrn.com/abstract=947640.

Natalie Glance, Matthew Hurst, Kamal Nigam, Matthew Siegler, Robert Stockton, and Takashi Tomokiyo. 2005. Deriving marketing intelligence from online discussion. In *Proc. KDD'05*, pages 419–428, New York, NY, USA. ACM.

Stephan Greene. 2007. *Spin: Lexical Semantics, Transitivity, and the Identification of Implicit Sentiment*. Ph.D. thesis, University of Maryland.

T. Groseclose and J. Milyo. 2005. A measure of media bias. *The Quarterly Journal of Economics*, 120:1191–1237.

B. Harden. 2006. On Puget Sound, It's Orca vs. Inc. *The Washington Post*. July 26, page A3.

Marti Hearst. 1992. Direction-based text interpretation as an information access refinement. In Paul Jacobs, editor, *Text-Based Intelligent Systems*, pages 257–274. Lawrence Erlbaum Associates.

Paul Hopper and Sandra Thompson. 1980. Transitivity in Grammar and Discourse. *Language*, 56:251–295.

E. Kako. 2006. Thematic role properties of subjects and objects. *Cognition*, 101(1):1–42, August.

Michael Laver, Kenneth Benoit, and John Garry. 2003. Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2):311–331.

M. Lemmens. 1998. *Lexical perspectives on transitivity and ergativity*. John Benjamins.

Beth Levin and Malka Rappaport Hovav. 2005. *Argument Realization*. Research Surveys in Linguistics. Cambridge University Press, New York.

Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago, IL.

Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. Which side are you on? identifying perspectives at the document and sentence levels. In *Proceedings of the Conference on Natural Language Learning (CoNLL)*.

Lanny W. Martin and Georg Vanberg. 2008. A robust transformation procedure for interpreting political text. *Political Analysis*, 16(1):93–100.

G. McKoon and T. MacFarland. 2000. Externally and internally caused change of state verbs. *Language*, pages 833–858.

M. Mulder, A. Nijholt, M. den Uyl, and P. Terpstra. 2004. A lexical grammatical implementation of affect. In *Proc. TSD-04, Lecture notes in computer science 3206*, pages 171–178). Springer-Verlag.

Tony Mullen and Robert Malouf. 2008. Taking sides: User classification for informal online political discourse. *Internet Research*, 18:177–190.

Mari Broman Olsen and Philip Resnik. 1997. Implicit Object Constructions and the (In)transitivity Continuum. In *33rd Proceedings of the Chicago Linguistic Society*, pages 327–336.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

James Pustejovsky. 1991. The Generative Lexicon. *Computational Linguistics*, 17(4):409–441.

Philip J. Stone. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proc. EMNLP*, pages 327–335.

Zhibao Wu and Martha Palmer. 1994. Verb Semantics and Lexical Selection. In *Proc. ACL*, pages 133–138, Las Cruces, New Mexico.