MORE THAN WORDS: SYNTACTIC PACKAGING AND IMPLICIT SENTIMENT

Greene & Resnik 2009

MOTIVATION for the TOPIC

- Businesses and organizations:
  Product, service and CRM benchmarking
  Market intelligence (product improvement)

- People:
  Finding opinions while purchasing product
  Finding opinions on political topics (trends)

- Advertisement: (a sub-component technology)
  Placing ads in the user-generated content
  Place an ad when one praises a product
  Place an ad from a competitor if one criticizes a product.

- Information Search & Retrieval:
  Providing general search for "opinions".

Sentiment Analysis – Expanding Resources

- **Lexicons**
  - General Inquirer (Stone et al., 1966)
  - OpinionFinder lexicon (Wiebe & Riloff, 2005)
  - SentiWordNet (Esuli & Sebastiani, 2006)

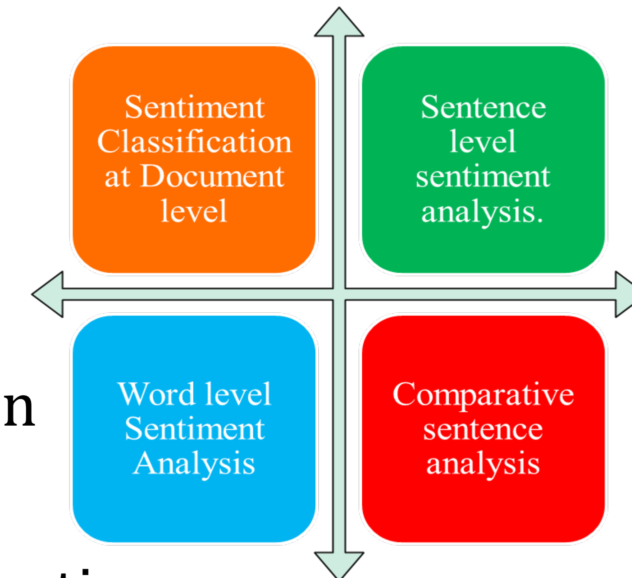
- **Annotated Corpora**
  - Used in statistical approaches (Hu & Liu 2004, Pang & Lee 2004)
  - MPQA corpus (Wiebe et. al, 2005)

- **Tools**
  - Algorithm based on minimum cuts (Pang & Lee, 2004)
  - OpinionFinder (Wiebe et. al, 2005)
  - Engines – Attensity, Lexalytics, IBM (??? 2009)

ACTIVE RESEARCH AREAS of INTEREST

There are primarily four different problems predominating sentiment detection in text research community, namely:

- Subjectivity classification
- Word sentiment classification
- Document sentiment classification
- Opinion extraction.



PROBLEM

How to interpret features for sentiment detection?

- Bag of words (IR)
- Annotated lexicons (WordNet, SentiWordNet)
- Syntactic patterns

Which features to use?

- Words (unigrams)
- Phrases/n-grams
- Sentences

CHALLENGES

How to interpret features for sentiment detection?

- Need to consider other features due to ….
  - Words alone may not convey true sentiment
    - Every time I read Pride and Prejudice I want to dig her up and beat her over the skull with her own shin-bone.
  - Subtlety of sentiment expression
    - irony
- Domain/context dependence
  - Words/phrases can mean different things in different contexts and domains

SOME BACKGROUND on the TOPIC

- Sentiment Analysis (SA) treated in detail in Pang & Lee

- Described in the literature under a number of names which are similar though not necessarily synonymous ……. subjectivity analysis, opinion analysis/extraction, sentiment mining, etc.

- Most initial SA work focused on lexical indicators, frequency counts, word polarity - Shallow Analysis

- Approaches that examine semantic properties of text have drawn increased interest in the field in recent years …… Resnik is one of a number of linguistic researchers who has presented work in this area

An INTRODUCTION to the PAPER

- Most work on analysis of people's attitudes relies on words that express overt opinions
- Underlying perspective can also reside in less obvious linguistic choices
- Language can be used "to select some aspects of a perceived reality and make them more salient in a communicating text; may promote a particular problem definition, moral evaluation or recommendation
- Entman calls this framing, and deliberately framing in a way that manipulates or deceives is referred to as spin

MOTIVATION FOR THE PAPER – read Greene 2007

- This paper describes an approach to this problem that focuses not on lexical indicators, but on the syntactic “packaging” of ideas thru implicit syntactic indicators
- The authors establish a strong predictive connection between linguistically well motivated features (sentence structure) and implicit sentiment (perspective)
- They demonstrate how computational approximations of these features can be used to improve sentiment classification results
- Not really a new idea - linguists have long studied syntactic variation in descriptions of the same event, often under the general heading of syntactic diathesis alternations (Levin, 1993 and others) and how elements of meaning are syntactically reflected

SOME PRIOR RESEARCH WORK

Classifying implicit sentiment is less studied in NLP literature – we have to look elsewhere

- Journalism studies (Gentzkow and Shapiro, 2006)
- Marketing and business intelligence (Glance, 2005)

Computational linguistics work in implicit sentiment

- Identification of perspective (Lin, et. al., 2006)
- Perspective (Martin and Vanberg, 2008)
- Uses features based on sentence logical form (Gamon, 2004) ….. Most closely related work
- Argument structure features with lexical information (Mulder, 2004)
- Predicting votes on floor debate speeches (Thomas, 2006)

THE PAPER IN A NUTSHELL

- Formulates a hypothesis about connection between sentence structure and implicit sentiment
- Attempts to validate the hypothesis by means of a human ratings study
- Introduces OPUS (observable properties for underlying semantics) method for approximating relevant semantic properties automatically as features in supervised learning
- Demonstrates how these features improve on the existing state of the art in automatic sentiment classification
- Contends that set of underlying components of meaning motivated by lexical semantics literature can be used as basis for statistical classifier models to predict sentiment

THE UNDERLYING HYPOTHESIS

- Speakers employ specific constructions in a manner that exploits these (sometimes subtle) differences in meaning in a way that reveals, intentionally or not, through properties, aspects of the speakers' perspective

- Conscious (or not) choice of grammatical framing is accomplished by grammatical structure

- Syntactic reflections of these properties can be exploited as features for text classification tasks, even in the absence of overt opinion

FRAMING MAKES A DIFFERENCE…

(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.

OR …….

- Consider:
  - Millions of people starved under Stalin (inchoative)
  - Stalin starved millions of people (transitive)
- The latter will be perceived as more negative toward Stalin, because the transitive syntactic frame tends to be connected with semantic properties such as intended action by the subject and change of state in the object
- “Kill verbs” provide particularly strong examples of such phenomena, because they exhibit a large set of semantic properties canonically associated with the transitive frame (Dowty, 1991).

IMPLICIT SENTIMENT – A PERSPECTIVE

Implicit Sentiment - differences in linguistic form indicate at least some difference in meaning (Bolinger, 1968). …….. Greene, 2007. PHd Dissertation.

- Include everything from classic diathesis alternations to differences in the nominal forms for discourse
- Distinguish between perspective and subjective/ objective detection
- Examples
  - My toy broke.
  - I broke my toy.

Why syntax?

- the words in the above sentences are the same but something is different
- difference: way words are put together … the structure of the sentence

LINGUISTIC MOTIVATION

Syntactic diathesis alternations (syntactic variation or packaging)

- Verbs can be used in different frames with slight differences in semantic meaning

	break	**climb**
causative	X broke Y	X climbed Y
inchoative	Y broke	Y climbed *

* Breaking event entails change of state in Y, climbing event does not

Dowty(1991) and Hopper & Thompson(1980)

- 13 semantic properties organized into 3 groups

X (subject)	verb (event/state)	Y (direct object)
volitional involvement in event or state causation of the event sentence awareness and/or perception causing a change of state in Y kinesis or movement existence independent of the event	defined endpoint punctuality	affectedness change of state (lack of kinesis) or movement (lack of kinesis) existence

Dowty's relevance here

Dowty's theory of "thematic proto-roles" is based on the premise that the surface expression of (verbs') arguments in linguistic expressions is closely connected to properties of those arguments and of the event.

- If the referent of an argument is volitional and causal with respect to the event communicated by the verb, properties traditionally associated with thematic role of agent, then more likely to surface in subject position.

- If the referent of the argument undergoes a change of state and is causally affected by another participant in the event, properties traditionally associated with a patient thematic role, then it is more likely to surface as an object.

- 1)"Israeli Troops Shoot Dead Palestinian in W. Bank" – volitional agent, result and object
- 2) "Israeli Girl Killed, Fueling Cycle of Violence" – omits the overt agent; no argument from which to infer volition

So, authors predict that the expression of sentiment is connected with how particular entities are profiled ………… revealed in text by the grammatical relations in which they appear

EMPIRICAL VALIDATION - Experiment #1

Accomplished by varying the syntactic form of event descriptions and showing that semantic properties predict the perceived sentiment – connection exists between syntactic choice and sentiment

Semantic property ratings

- Stimuli with 11 verbs of killing separated into 2 classes/ paradigms
  - transitive (externally caused) …. kill, shoot, assassinate, poison – profiles properties of the agent
  - ergative (internally caused) …. strangle, smother, choke, drown – profiles properties of the patient

- Two forms of syntactic description
  - transitive form …… The gunmen shot the opposition leader.
  - nominalized form ….. The shooting killed the opposition leader.

- 18 participants (native speakers) rated sentences on a scale of 1 to 7
  - i.e. "In this event, how likely is it that (subject) chose to be involved?" where (subject) was *the gunmen* and *the shooting*

EMPIRICAL VALIDATION – Experiment #2

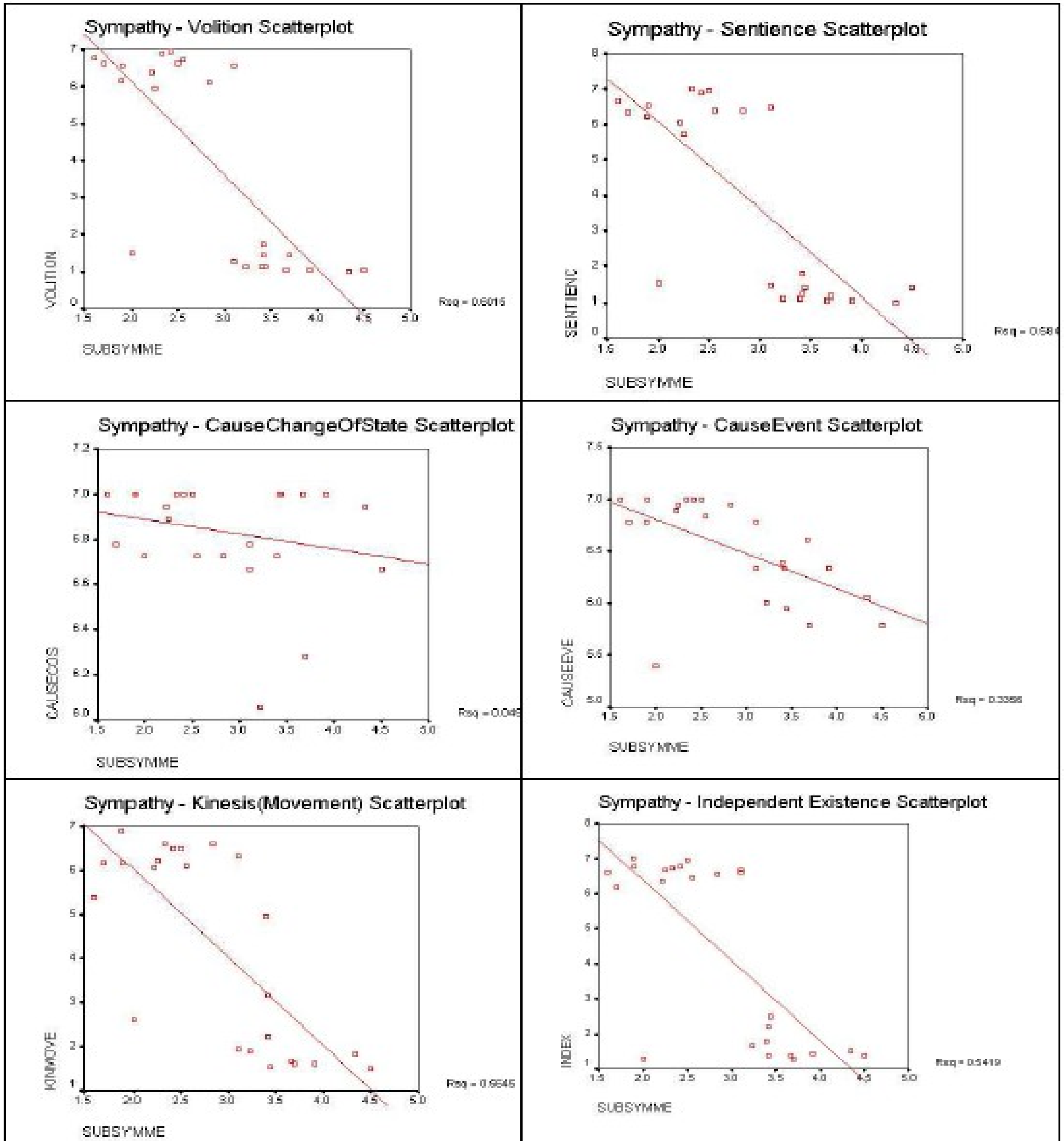
- newspaper-like paragraphs were constructed from materials
  - A man has been charged for the suffocation of a woman early Tuesday morning. City police say the man suffocated the 24-year-old woman using a plastic garbage bag. The woman, who police say had a previous relationship with her attacker, was 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.

- three alternative headlines presented
  - (a) Man suffocates 24-year old woman (transitive)
  - (b) Suffocation kills 24-year-old woman (nominalized subject)
  - (c) 24-year-old woman is suffocated (passive)

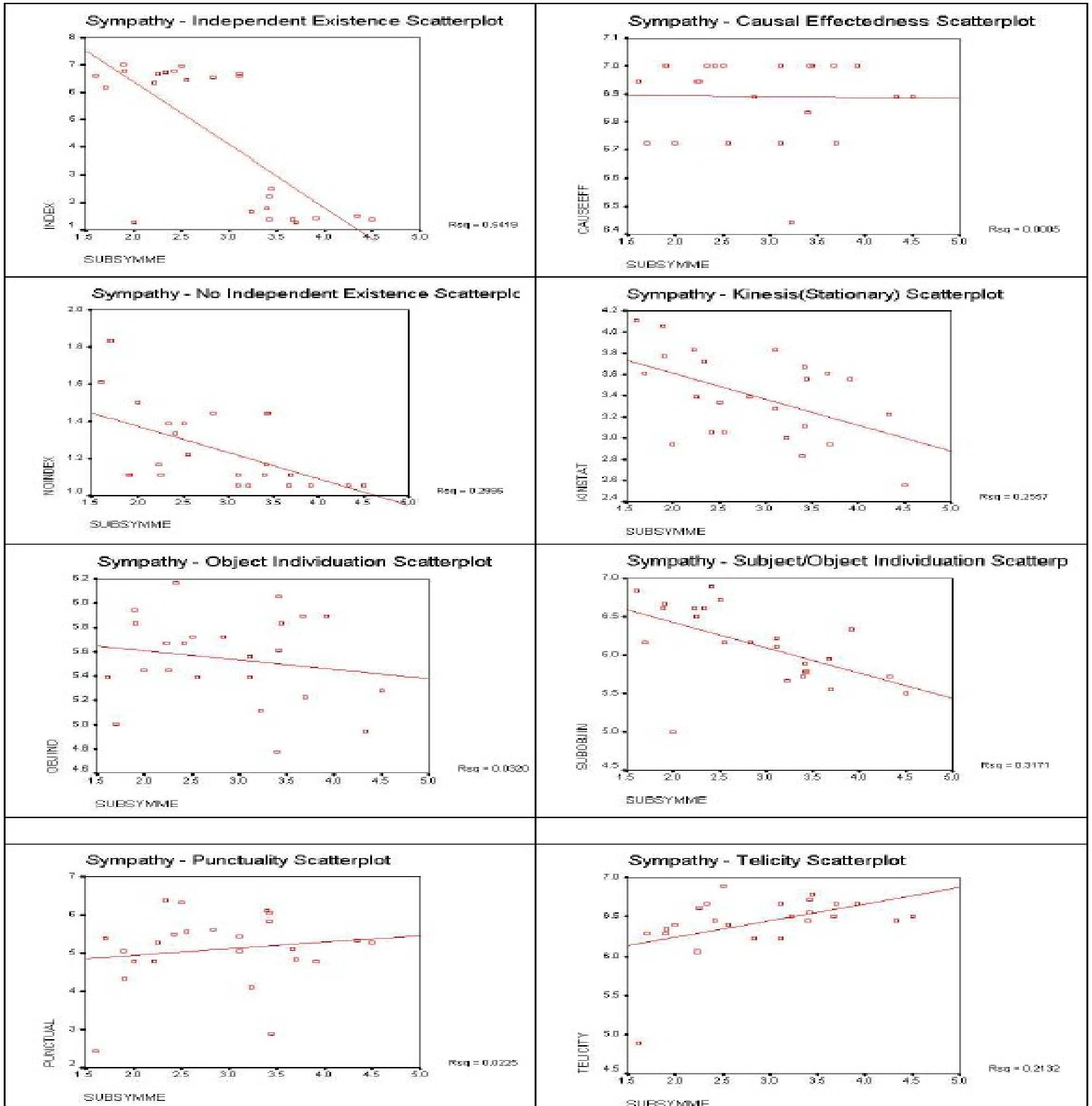
- 31 participants (native speakers) rated headlines on a scale of 1 to 7
  - i.e. "How sympathetic or unsympathetic is this headline to the perpetrator?"

The sympathy rating here is deemed a measure of sentiment …

SCATTERPLOTS – IV by DV (SYMPATHY)



SCATTERPLOTS – IV by DV (SYMPATHY)



Regression model with Volition, Telicity, and Verb as predictors

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.880(a)	.775	.741	.42195	1.896

a Predictors: (Constant), VERBITEM, VOLITION, TELICITY
b Dependent Variable: SYMPATHY

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12.260	3	4.087	22.953	.000(a)
	Residual	3.561	20	.178		
	Total	15.821	23			

a Predictors: (Constant), VERBITEM, VOLITION, TELICITY
b Dependent Variable: SYMPATHY

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.517	1.686		.307	.762
	VOLITION	-.200	.036	-.653	-5.594	.000
	TELICITY	.589	.257	.271	2.290	.033
	VERBITEM	-.105	.029	-.392	-3.606	.002

a Dependent Variable: SYMPATHY

EMPIRICAL VALIDATION

- **Results**
  - significant effect of syntactic form on sympathy toward perpetrator
    - linear mixed model ANOVA; $F(2, 369) = 33.902$, $p < .001$
      - Recall: p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true
    - transitive form of headline – significant lower sympathy
    - MLR models used to establish the degree to which specific semantic components of transitivity in a clause predict the IS attributable to clause

regression analysis (relationship between syntactic form and sentiment)

- independent variables – the 13 semantic property ratings plus verb identity
- dependent variable – sympathy rating
- sympathy negatively correlated with: volition ($r = -.776$)
sentience($r = -.764$)
kinesis(movement) ($r = -.751$)

" ….. this ratings study confirms the influence of syntactic choices on perceptions of implicit sentiment." (pg. 506)

- multiple regression
  - independent variables – verb, volition, telicity(defined endpoint)
  - $R = .88$, $R^2 = .78$ ($p < .001$)
  - When adjusted to account for small # of observations, $R^2 = .741$, that is the model accounts for ~75% of the variation in sympathy ratings

EXPERIMENT PURPOSE

- Authors propose (a) syntactic framing involves manipulation of semantic properties and (b) there is relation between syntactic choices and implicit sentiment
- But the point is that sympathy towards the subject increases with (if variables are assumed independent) non-volitionality, non-animacy, non-kinesis or (if not) volition and telicity

So, the hypothesis that the form of the event encoding affects sentiment about the subject in the event encoding was confirmed – <u>transitive surface encoding predicts less perceived sympathy for the perpetrator</u>

OBSERVABLE APPROXIMATION

BUT, properties like volition, telicity, etc. are not directly observable in nature and automated annotators and labeled training data doesn't readily exist. What do we do?

Enter **OPUS** (observable proxies for underlying semantics) - the linguistically motivated features derived and used in classification

A compromise between construction level syntactic distinctions (described in prior section) and annotation of fine-grained semantic properties

Key Idea:

- use observable grammatical relations from usage of terms relevant to that particular domain of interest as proxies for the underlying semantic properties that gave rise to the syntactic relations
- then automatically created features based on the proxies can be used in creating a classification scheme (described later)

OBSERVABLE APPROXIMATION (2)

OPUS

- Means of approximating the relevant semantic properties as features in a supervised learning setting
- Use observable grammatically relevant features as proxies for the underlying semantic properties
- There is a set of T of terms relevant to a collection; t is a term in T

$R(t) = R^{t}_{domain} / R^{t}_{reference\ (BNC)}$ relative frequency ratio – defined parameter

where $R^{t}_{c} = f^{t}_{c} / N_{c}$

the ratio of term t's frequency in corpus c to size N_c of that corpus

- R(t) is measure of term's prevalence in a collection relative to the corpus
- OPUS features are syntactic dependency relations involving terms in T
- Added features: TRANS:v (transitive) and NOOBJ:v (no direct object)

PRODUCING OPUS FEATURES from a DOCUMENT

- Requires syntactic as well as semantic analysis

- Requires a parser to extract document features representing grammatical relations

- Extract typed dependencies (subj, obj, indobj)

- The extraction method avoids data sparseness issues that occur in the use of full relational triples

OPUS FEATURE EXTRACTION - Example

Sentence:

Life Without Parole does not eliminate the risk that **the prisoner will murder a guard, a visitor, or another inmate.**

Constituent Parse (excerpt):

```
  (S
    (NP (DT the) (NN prisoner))
    (VP (MD will)
      (VP (VB murder)
        (NP
          (NP (DT a) (NN guard))
          (, ,)
```

Grammatical relations (excerpt):

```
nsubj(murder, prisoner)
aux(murder, will)
dobj(murder, guard)
```

OPUS features (excerpt):

```
TRANS-murder
murder-nsubj
nsubj-prisoner
murder-aux
aux-will
murder-dobj
dobj-guard
```

OPUS – SOME CAVEATS

Level of semantic analysis required to understand how the semantic components of transitivity are actually reflected in any given clause is well beyond the current capabilities of natural language processing systems. ………………… so, can OPUS be implemented practically?

There is an assumption here that consistent use of a language processing system will at least be consistent and systematic in its errors. In that context, machine learning techniques can effectively identify and exploit features amidst the noise. (See Gamon, 2004 for a treatment of this)

The key, therefore, is to identify features that are observable reflexes of the semantic components that can be practically extracted by current NLP techniques, even if noisily so.

COMPUTATIONAL APPLICATION

- **Two studies demonstrating the value of semantic features in sentiment classification and improvements on existing work**
  - (1) predicting opinions of the death penalty
    - corpus comprising documents from 5 pro- and 3 anti-death penalty sites (596 documents per each side)
    - N most frequent stemmed bigrams used as baseline feature set, where N = # of OPUS features used in the comparison condition
    - OPUS features created for:
      - 14 kill verbs (total of N = 1016 distinct features)
      - 117 verbs for which the relative frequency ratio was > 1.0 (total of N = 7552 distinct features) ; verbs describing physical force
    - OPUS features provide substantial and statistically significant gains (p < .001)

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

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

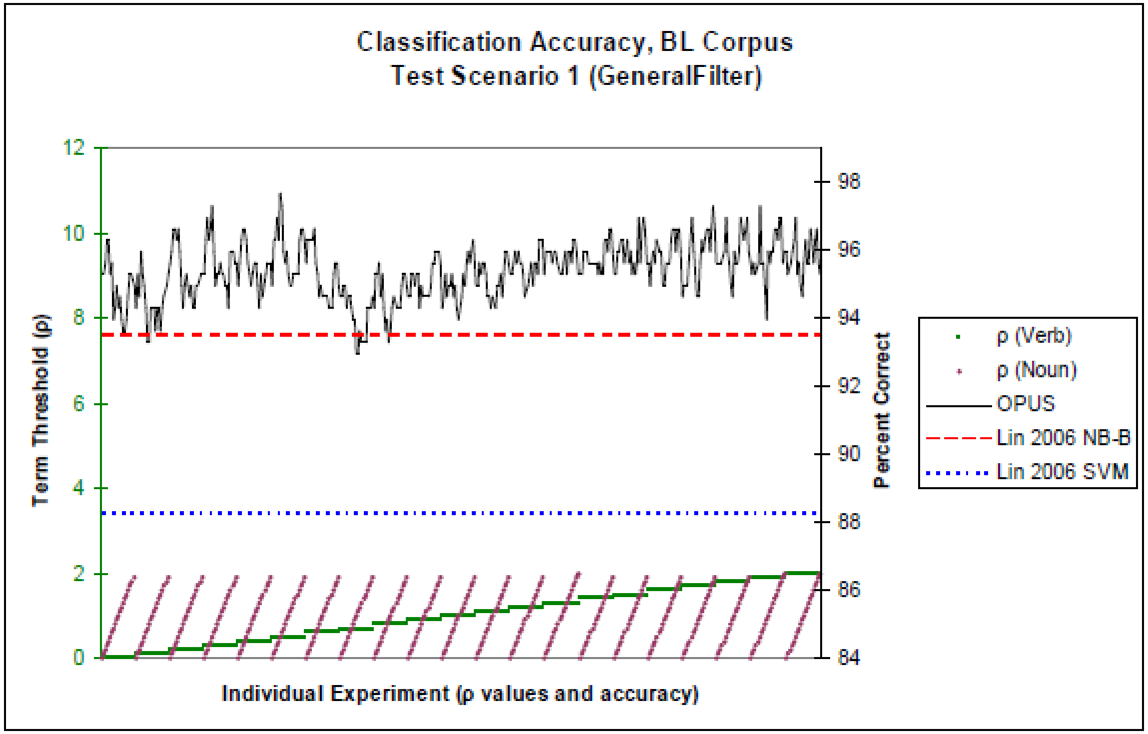
But DP corpus exhibits some uniformity in lines of argumentation. To test generalization, try a more diverse corpus ....

COMPUTATIONAL APPLICATION

- **Two studies demonstrating the value of semantic features in sentiment classification**
  - (2) predicting points of view in the Israeli-Palestinian conflict
    - Bitter Lemons corpus comprising opposing viewpoints on controversial issues
    - total of 297 documents from each viewpoint, averaging 700-800 words in length
    - authors use the WEKA SVM classifier and compare their results to the SVM and naïve Bayes classifier with full Bayesian inference of Lin et. Al. (2006)
    - OPUS features on terms for which $\log(R(t)) > p$, where p is a threshold for the relative frequency ratio
    - Two test scenarios:
      - TS1: documents written by site's guests as training data: documents from site's editors as test data
      - TS2: documents from site editors as training data; documents written by site's guests as test data

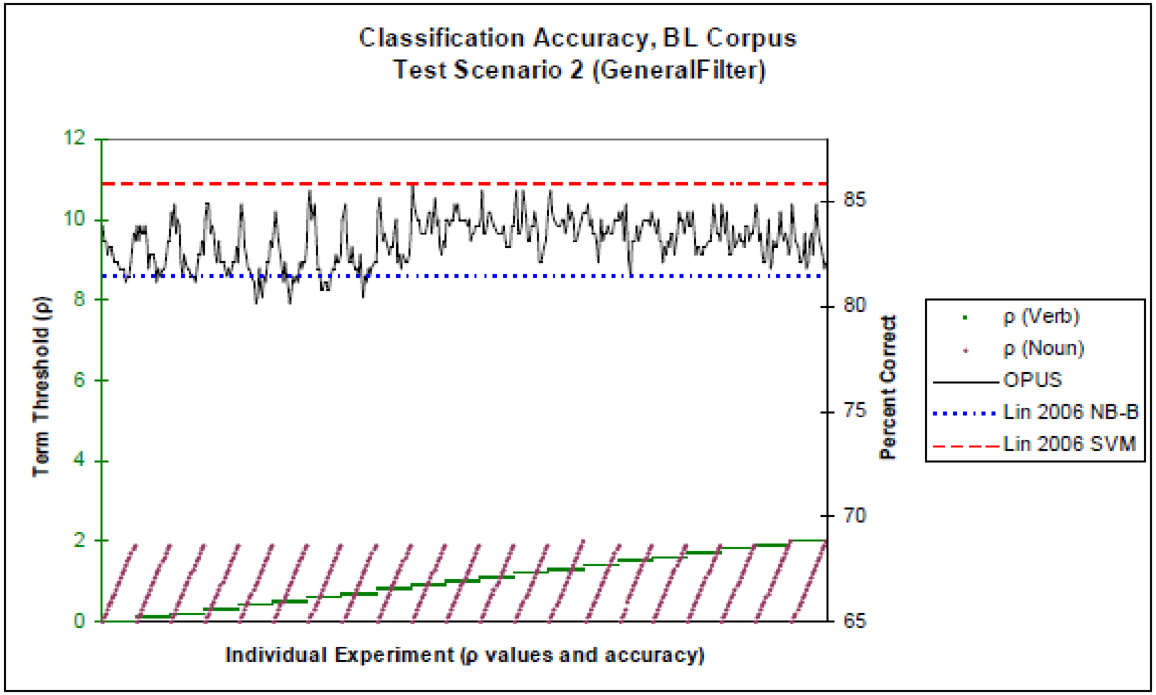
COMPUTATIONAL APPLICATION

- **Two studies demonstrating the value of semantic features in sentiment classification**
  - (2) predicting points of view in the Israeli-Palestinian conflict
    - Test Scenario 1 results:
      - Greene & Resnik: SVM had an average accuracy of 95.41% (ranging from 92.93% to 97.64%)
      - Lin et. al.: SVM with 88.22% accuracy and NB-B classifier with 93.46% accuracy (constant across all variation for both)



COMPUTATIONAL APPLICATION

- **Two studies demonstrating the value of semantic features in sentiment classification**
  - (2) predicting points of view in the Israeli-Palestinian conflict
    - Test Scenario 2 results:
      - Greene & Resnik: SVM had an average accuracy of 83.12% (with a maximum of 85.86%)
      - Lin et. al.: SVM with 81.48% accuracy and NB-B classifier with 85.85% accuracy (constant across all variation for both)
      - Performed worse than TS1 because editor-authored documents make up a smaller training set



Overall, positive results show that these OPUS features “may also improve conventional opinion labeling for subjective text.” (pg. 510)

CRITICISMS OF THE STUDY

- Several of the corpora used (e.g. the Bitter Lemons corpus among others) exhibit more pronounced sentiment examples in general ……….. using a corpus displaying more subtle examples of sentiment might have provided a better test of sentiment detection
- The utility of the features w/r to labeling of subjective text is more conjecture at this point ………… it requires additional study
- As topics get more diverse, and are less focused, the classification accuracy falls off – indicates that the technique doesn't generalize well across domains – the authors omit dis-confirming information from the paper
- There are better methods than relative frequency to identify candidate features

CONCLUSION

Criticisms aside, this is an important work.

- Paper presents evidence for the role of semantic properties in sentiment judgments
- Work provides an explicit and empirically supported connection between theoretically motivated work in lexical semantics and readers' perception of sentiment
- Reported positive sentiment classification results within a standard supervised learning setting, employing a practical first approximation to those semantic properties
- Introduced OPUS features as a possible means of improving sentiment labeling for subjective text

A BRIEF WORD ON PROJECT

- New resources have been conceived/described, and in some cases built to aid in the classification of sentiment in documents
- Specifically, Italian researchers have created resource known as SentiWordNet, which is modeled on the WordNet lexicon
- Tools including SentiFul and SentiFrameNet have also been described and proposed in the literature, though not all have working prototypes that have been widely disseminated and made available to researchers
- Project endeavors to either utilize available resources or construct working copies of the described resources based on research descriptions and test them against sentiment-driven text from several domains to determine their current utility
- Stretch goal is to gain insight into the construction of such lexicons and provide insight into possible enhancements in future releases to make them more robust for general use or more domain-specific

THE DATASETS – Sentiment–Annotated Data is Difficult to Procure in Any Domain

- MPQA Corpus – Univ. of Pittsburgh: This corpus contains news articles and other text documents manually annotated for sentiment and other private states.
- EDGAR (SEC) and CRSP (Center for Research on Security Prices) data: 10K and MD&A company data from 1994-2008; often used as a proxy for forward security pricing; Annotated version received from University of Notre Dame researchers based on liability research paper of 2009; Sentiment confirmation via OpinionFinder.
- Legal blog annotated data from 2009 based on collection created by running a set of directed queries against an assortment of Web search engines focused on legal blogs. Run produced roughly 200 blog entries consisting of approximately 1,000 sentences. Collection focused on the original blog entries, not on subsequent responses. Courtesy of Thomson Legal and Regulatory, St. Paul, Minnesota, Study in Proceedings of the 11th International Conference on Artificial Intelligence and Law (ICAIL09).