Computing Attitude and Affect in Text: Theory and Applications

THE INFORMATION RETRIEVAL SERIES

Series Editor:
W. Bruce Croft
University of Massachusetts, Amherst

Also in the Series:

INFORMATION RETRIEVAL SYSTEMS: ***Theory and Implementation,*** *by Gerald Kowalski;* ISBN: 0-7923-9926-9

CROSS-LANGUAGE INFORMATION RETRIEVAL, *edited by Gregory Grefenstette;* ISBN: 0-7923-8122-X

TEXT RETRIEVAL AND FILTERING: ***Analytic Models of Performance***, *by Robert M. Losee;* ISBN: 0-7923-8177-7

INFORMATION RETRIEVAL: UNCERTAINTY AND LOGICS: ***Advanced Models for the Representation and Retrieval of Information***, *by Fabio Crestani, Mounia Lalmas, and Cornelis Joost van Rijsbergen;* ISBN: 0-7923-8302-8

DOCUMENT COMPUTING: ***Technologies for Managing Electronic Document Collections***, *by Ross Wilkinson, Timothy Arnold-Moore, Michael Fuller, Ron Sacks-Davis, James Thom, and Justin Zobel;* ISBN: 0-7923-8357-5

AUTOMATIC INDEXING AND ABSTRACTING OF DOCUMENT TEXTS, *by Marie-Francine Moens*; ISBN 0-7923-7793-1

ADVANCES IN INFORMATIONAL RETRIEVAL: ***Recent Research from the Center for Intelligent Information Retrieval***, *by W. Bruce Croft*; ISBN 0-7923-7812-1

INFORMATION RETRIEVAL SYSTEMS: ***Theory and Implementation,*** **Second Edition**, *by Gerald J. Kowalski and Mark T. Maybury;* ISBN: 0-7923-7924-1

PERSPECTIVES ON CONTENT-BASED MULTIMEDIA SYSTEMS, *by Jian Kang Wu; Mohan S. Kankanhalli;Joo-Hwee Lim;Dezhong Hong;* ISBN: 0-7923-7944-6

MINING THE WORLD WIDE WEB: ***An Information Search Approach***, *by George Chang, Marcus J. Healey, James A. M. McHugh, Jason T. L. Wang;* ISBN: 0-7923-7349-9

INTEGRATED REGION-BASED IMAGE RETRIEVAL, *by James Z. Wang;* ISBN: 0-7923-7350-2

TOPIC DETECTION AND TRACKING: Event-based Information Organization, *edited by James Allan;* ISBN: 0-7923-7664-1

LANGUAGE MODELING FOR INFORMATION RETRIEVAL, *edited by W. Bruce Croft; John Lafferty;* ISBN: 1-4020-1216-0

MACHINE LEARNING AND STATISTICAL MODELING APPROACHES TO IMAGE RETRIEVAL, *by Yixin Chen, Jia Li and James Z. Wang;* ISBN: 1-4020-8034-4

INFORMATION RETRIEVAL: ***Algorithms and Heuristics,*** *by David A. Grossman and Ophir Frieder,* 2nd ed.; ISBN: 1-4020-3003-7; PB: ISBN: 1-4020-3004-5

Computing Attitude and Affect in Text: Theory and Applications

Edited by

James G. Shanahan

Clairvoyance Cooperation,
Pittsburgh, PA, U.S.A.

Yan Qu

Clairvoyance Cooperation,
Pittsburgh, PA, U.S.A.

and

Janyce Wiebe

University of Pittsburgh,
PA, U.S.A.

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-4026-1 (HB)
ISBN-13 978-1-4020-4026-9 (HB)
ISBN-10 1-4020-4102-0 (e-book)
ISBN-13 978-1-4020-4102-0 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

Printed in the Netherlands.

TABLE OF CONTENTS

Preface

Human Language Technology (HLT) and Natural Language Processing (NLP) systems have typically focused on the "factual" aspect of content analysis. Other aspects, including pragmatics, opinion, and style, have received much less attention. However, to achieve an adequate understanding of a text, these aspects cannot be ignored.

The chapters in this book address the aspect of subjective opinion, which includes identifying different points of view, identifying different emotive dimensions, and classifying text by opinion. Various conceptual models and computational methods are presented. The models explored in this book include the following: distinguishing attitudes from simple factual assertions; distinguishing between the author's reports from reports of other people's opinions; and distinguishing between explicitly and implicitly stated attitudes. In addition, many applications are described that promise to benefit from the ability to understand attitudes and affect, including indexing and retrieval of documents by opinion; automatic question answering about opinions; analysis of sentiment in the media and in discussion groups about consumer products, political issues, etc.; brand and reputation management; discovering and predicting consumer and voting trends; analyzing client discourse in therapy and counseling; determining relations between scientific texts by finding reasons for citations; generating more appropriate texts and making agents more believable; and creating writers' aids. The studies reported here are carried out on different languages such as English, French, Japanese, and Portuguese.

Difficult challenges remain, however. It can be argued that analyzing attitude and affect in text is an "NLP"-complete problem. The interpretation of attitude and affect depends on audience, context, and world knowledge. In addition, there is much yet to learn about the psychological and biological relationships between emotion and language.

To continue to progress in this area in NLP, more comprehensive theories of emotion, attitude and opinion are needed, as are lexicons of affective terms and knowledge of how such terms are used in context, and annotated corpora for training and evaluation.

This book is a first foray into this area; it grew out of a symposium on this topic that took place at Stanford University in March, 2004, under support from American Association for Artificial Intelligence (AAAI). Several of the presentations were extended into the chapters that appear here. The chapters in this collection reflect the majors themes of the workshop, corresponding to a balance among conceptual models, computational methods, and applications. The chapters in this book are organized along these themes into three broad, overlapping parts.

Linguistic and Cognitive Models

The chapters in the first part of this book explore linguistic and cognitive models which could support developing richer computational models of attitude and affect. This section begins with Polanyi and Zaenen's fascinating study of attitudinal valence (or polarity) as it is expressed in context. While individual words often suggest a negative or positive attitude, such as "horrible" and "great", respectively, the context of a word may change its base valence. Polanyi and Zaenen describe and illustrate a number of such contextual valence shifters, both intra-sentential (e.g., negatives and modals) and inter-sentential (e.g., discourse connectives and multi-entity evaluation).

The next chapter, by Bergler, also explores linguistic devices for conveying attitudes in text, namely reported speech expressions which convey attitudes. Bergler argues that reported speech serves to segment information into discourse segments called profiles. Each profile involves such things as degree of credibility of the source of the information, and the role the source has in the argumentative structure of the text. Bergler performs a detailed profile analysis of an extended story, and discusses extending this type of analysis to other attributes than reported speech.

Like Polaryi and Zaenen, and Bergler, Karlgren et al. focus on contextual aspects of linguistic expressions of attitudes. Their particular objects of study are attitude expressions which are internally structured. They argue that simple lists of attitudinal terms are not sufficient for recognizing attitudes in texts: it is often only in particular lexical and syntactic patterns that words convey attitudes. They present interesting results of a corpus study suggesting that certain syntactic contexts are more likely to be loci of attitudes, and that this is realized in stylistic differences between opinionated text types such as editorials and more objective text types such as reporting news articles.

The second set of chapters in this section address cognitive as well as linguistic issues in understanding attitude and affect in text. Green's chapter presents the results of a qualitative analysis of letters written by genetic counselors to their clients. The goal is to find stylistic features that would be salient for natural generation systems in this genre. Her study suggests that perspective must be taken into account to generate stylistically appropriate text. Green identifies a number of perspectives in this genre, including specific agents such as the author and client, as well as abstract perspectives such as education and research. As a generation system assumes different perspectives while generating such a letter, it should choose forms of reference, tenses, types of evidential language, and so forth to reflect that perspective.

Morris and Hirst's chapter addresses readers' perceptions of lexical semantic relations, in particular the perceived subjectivity of such relations. They perform a study to assess the degree of individual differences in readers' interpretations of lexical chains, which are groups of related words that create lexical cohesion. The results showed that subjects identified a common core of groups of related words in text, but also exhibited individual differences. Such knowledge could help NLP systems recognize which types of text meaning can be expected to be shared by most readers, and understand and generate text appropriately.

Bucci and Maskit's presentation of a "Weighted Referential Activity Dictionary" is the most psychologically oriented chapter in the collection. Bucci and Maskit use their dictionary in computer modeling of a psycholinguistic variable which they call Referential Activity (RA). RA ratings measure the degree to which language connects to nonverbal experiences such as bodily and emotional experience. In Bucci and Maskit's model, RA is mainly indicated by domain-independent stylistic attributes of language, aspects of which are included in their dictionary. The chapter presents compelling RA analyses of literary passages, and describes a method for assigning RA weights to dictionary entries. Their study reveals differential linguistic roles of particular lexical items in producing vivid versus abstract texts. They plan to investigate the psychological significance of these differences in future work.

The final two chapters in this part of the book present annotation schemes, i.e., schemes for manually labeling texts to create data for training and evaluating NLP systems. The chapter by Rubin et al. presents a framework for coding the writer's certainty in text. They categorize a set of linguistic certainty markers (such as "probably"and "allegedly") along four dimensions – level (degree of certainty), perspective (whose certainty is being encoded), focus (abstract versus factual information), and time (past, present, or future). They perform an empirical study of their framework in which they applied their annotation scheme to 32 newspaper articles. Among their findings are that editorials contain more explicit certainty markers than news articles, and that a few specific combinations of dimension values dominate in editorials. The framework and empirical results will be informative for developing automatic certainty identification systems.

The chapter by Stoyanov et al. addresses annotations for Multi-Perspective Question Answering (MPQA), whereby an NLP system answers opinion-oriented questions. To be successful, an MPQA system will presumably need to recognize and organize the opinions expressed in one or more documents. An annotation scheme for encoding such opinions has been developed and evaluated in previous work. This chapter investigates the utility of that annotation scheme for MPQA processing. It first describes a new corpus of multi-perspective questions and answers. It then presents the results of a study investigating the usefulness of the earlier opinion annotations for multi-perspective versus fact-based question answering. Their findings are that opinion annotations can be useful for MPQA if used appropriately.

Lexical Resources and Attitude/Affect Recognition and Generation

The first two chapters in this part of the book focus on lexical resources that could support recognition and generation of attitude and affect. Lexicons of words of emotion-conveying potential have been used in much work for identifying and generating affect. The chapter by Grefenstette et al. addresses the problems of automatically extracting affect words for expanding the coverage of existing affect lexicons, and of automatically assigning the affect words along multiple semantic axes. Emotive patterns are used as seeds to extract affect words from the Web. Through evaluation of the precision and recall of the extracted words, the authors show that it is possible to identify lexical patterns for finding emotion-bearing affect words with high precision. Once the affect words are extracted, the authors discuss ways to automatically assign the words along the different semantic axes using measures similar to point-wise mutual information. The measures show promise for finding degrees of belongingness to the semantic classes while at the same time assigning degrees of intensity to the affect words.

In the following chapter, Mathieu first presents a manually constructed lexicon of French verbs of emotion with positive, negative, or neutral affect. Thirty-three semantic classes are proposed and the classes are arranged in graphical structures through links of intensity and antonyms. French verbs of each class are described by simple attribute-pair type properties such as whether a verb accepts a non-agentive subject or not. The lexicon is evaluated for identifying positive, negative, or neutral affect of sentences from French Letters to the Editors texts. The evaluation shows that taking into account the intensity of verbs of emotion produces better classification results.

The next three chapters present computational methods for recognizing attitude and affect in text. The first chapter by Bethard et al. addresses the tasks of detecting propositional opinions and detecting holders of these opinions. Unlike a variety of previous work on separating facts from opinions at the document or sentence level, this paper focuses on determining the opinion status of a smaller piece of text. Propositional opinions are opinions that are generally found as the sentential complements of a predicate. The authors use supervised statistical classification methods for proposition detection and opinion-holder detection, incorporating semantic constituent labeling, opinion-oriented words, and syntactic features such as the presence of complex adjective phrases.

The next chapter by Chambers et al. presents approaches for automatically tagging the attitude of the speakers in transcribed dialogues. The authors explore several n-gram- and vector-based approaches and present results in a marriage-counseling domain and the Switchboard Corpus. In the marriage-counseling domain, each transcript is broken into thought units that are manually annotated with tags classifying the attribute and emotional commitment of the participants to a particular topic of discussion. The Switchboard corpus consists of conversations of random topics and has a richer tagging scheme. The performance results over both corpora are comparable, and the simple n-gram based approaches outperform or perform as well as the vector-based approaches. The authors also describe a Java tool for tagging attitude and affect which integrates the automatic classification capability.

The chapter by Teufel addresses the problem of automatically classifying academic citations in scientific articles according to author affect. The two rhetorical roles for citation analysis that are associated with affect in text include Contrast (comparison with, criticism of, or contract to other work) and Basis (agreement with or continuation of other work). Teufel examines discourse features such as section structure, history to classify author affect, in addition to other features such as semantic class of main verb, indicator phrases, etc. Such analysis aims at improving citation indexing through better detection of subjectivity in scientific text.

The last two chapters in this part of the book explore attitude and affect in text generation and summarization. Roman et al. explore the influence of affect and attitude on summarizing dialogues. In particular, they address the question of whether politeness should be reported in dialogue summaries and, if so, how politeness is reported. The chapter presents empirical studies designed to gather information about how people summarize dialogues. In these studies, a collection of four dialogues, involving a customer and vendor about buying a car, was used. Each dialogue was generated by an automated system with the politeness of the dialogue participants manipulated. Subjects were asked to summarize the dialogues from a particular dialogue participant's point of view. The studies showed that the percentage of summaries reporting some behavior information was higher when the dialogues were more impolite. This result is independent of the point of view and summary size. The studies also indicated that the point of view adopted by the summarizer biases the reporting of behaviors in their summaries. In particular, negative reporting of behavior information depends on the point of view of the summarizer rather than on the actual dialogue behavior. Tentative evidence showed that positive reporting is less subject to such bias.

While Roman et al. study how people's points of view influence human generation of text and summaries, Inkpen et al. explore a way of producing text with different attitudinal nuances by varying word choices. In particular, they examine nuances that differentiate near-synonyms relating to expressed attitude and text, and propose to transform the semantic orientation of a text automatically by choosing near-synonyms accordingly. The transformation of semantic orientation involves first representing text as an inter-lingual representation and a set of lexical nuances, and then replacing the words with attitudinal nuances in the original text by their near-synonyms according to the desired nuances.

Applications

The third part of this book focuses on applications of attitude and affect. The first two chapters in this section explore the categorization of text based on the manner in which a document is written rather than its content. In particular, both chapters use a computational model based on different aspects of systemic functional linguistic (SFL) theory. Whitelaw et al. present a study that demonstrates that the pronominal and determination systems of SFL are indeed powerful ways of characterizing interpersonal distance (between author and reader). They show empirically that this characterization of text is a robust means of recognizing financial scam email from regular email with a performance accuracy of 98% using a variety of machine learning algorithms. In contrast, Argamon and Dodick focus on conjunction, modality, and comment subsystems of SFL for genre-based text categorization of scientific articles in the historical and experimental sciences. Using a support vector machine trained on a systemic functional feature set (with no domain specific terms), they achieved over 83% accuracy for classifying articles according to field. The most highly-weighted features for each were consistent with hypothesized methodological differences between historical and experimental sciences.

The next two chapters in this section deal with applications that analyze the rhetorical structures of scientific papers. The first chapter by Feltrim et al. describes a system that uses argumentative zoning, a technique for identifying the rhetorical structure of text, as a thesis writing aid for graduate students working in Portuguese. Argumentative zoning techniques assign a label (drawn from possible rhetorical role labels such as background, purpose, results, and conclusion) to each sentence, indicating its argumentative role in a portion of text. The argumentative zoning algorithm (realized through a Naïve Bayes classier) is used to label each sentence as being a background, gap, purpose, methodology, result, conclusion, or outline. These rhetorical labels are then used by a rule-based system to identify problems in scientific text abstracts. A reported user study highlights the value of such a system for masters-level students. The work represents a successful adaptation of the argumentative zoning technique to the Portuguese language. In contrast, Di Marco et al., in their chapter, empirically validate a hypothesis that the use of hedges (words that make text more or less vague) is highly correlated with sentences that contain or surround citations. This study is based upon 985 peer-reviewed recent biology journal articles from the BioMed Central corpus. In addition, Di Marco et al. describe a system for classifying citations into 35 categories using a hand-built decision tree over cue-words, polarity switching words, and knowledge of the discourse structure of the article, among other features. Citation categories vary depending on the function of the citation, e.g., support or contrast.

The next two chapters in this section focus on aggregation of opinion from multiple sources. Nigam and Hurst describe an interesting polarity classifier which uses shallow NLP techniques and a topic-based classifier. They propose using a Bayesian statistical approach to aggregate the opinions expressed about a specific topic in Internet forums. Tong and Yager explore aggregating and characterizing opinion over time. They first create a time series of the subjects, opinions, and attitudes expressed in Internet sources. Subsequently, they generate linguistic summaries, using fuzzy set theory, which provide perspicuous overviews of the opinions expressed towards an event over a period of time.

The final three chapters of this section focus on empirical studies of deploying opinion-based systems. Koppel and Shtrimberg examine the use of sentiment analysis as a means of predicting future stock prices. Though their findings highlight that this is not a useful investment strategy, one potentially useful outcome of their work is a method for collecting labeled data for sentiment analysis, where data is labeled based upon the direction of relative large changes in stock price.

Salvetti, Reichenbach and Lewis describe an approach to opinion classification of movie reviews based upon feature selection (using part of speech tags), feature generalization (in terms of synonymy and hypernymy), and probabilistic classifiers (namely Naïve Bayes and Markov

Models). They note that using a simple thresholding of the log odds ratio of the positive and negative posterior probabilities can dramatically improve performance.

The final chapter, by Seki, Eguchi and Kando, focuses on multi-document summarization based on a topic/query and investigates the impact of using sentence and document-level genre information on building three types of summaries: summaries that concentrate on facts (events), opinions, or knowledge (definitions), respectively. The topic, characterized as a query, is used to retrieve/select documents from a collection of documents. The retrieved documents are then summarized using a clustering-based approach, where clusters and sentences within clusters are ranked and selected based upon similarity to the topic. The user is further allowed to select the type of summary required. The reported results on Japanese newswire articles show significant improvement in summary coverage and precision when combining sentence-level typing and genre classification information over baseline multi-document summarization techniques.

Target Audience

The book is intended for advanced undergraduate and graduate students, as well as a broad audience of professionals and researchers in computer science, engineering, information science, and content analysis who have an interest in the subjective aspects of text. The subject matter in this book is far ranging, including conceptual models, computational models, and applications.

Acknowledgements

We thank Helen Perilioux, Jesse Montogomery, and Debbie Moran for their help in editing this book.

James G. Shanahan
Yan Qu
Janyce Wiebe
Pittsburgh, June 2005

Chapter 1

Contextual Valence Shifters

Livia Polanyi
FXPAL
3400 Hillview Ave, Bldg. 4
Palo Alto CA 94304
E-mail: polanyi@fxpal.com

Annie Zaenen
PARC,
3333 Coyote Hill Road,
Palo Alto, CA 94304 USA
E-mail: zaenen@parc.com

Abstract

In addition to describing facts and events, texts often communicate information about the attitude of the writer or various participants towards material being described. The most salient clues about attitude are provided by the lexical choice of the writer but, as discussed below, the organization of the text also contributes information relevant to assessing attitude. We argue that the current work in this area that concentrates mainly on the negative or positive attitude communicated by individual terms (Edmonds and Hirst, 2002; Hatzivassiloglou and McKeown, 1997; Turney and Littman, 2002; Wiebe et al., 2001) is incomplete and often gives the wrong results when implemented directly. We then describe how the base attitudinal valence of a lexical item is modified by lexical and discourse context and propose a simple, "proof of concept" implementation for some contextual shifters.

Keywords: attitude, discourse, valence shifters, genre structure, multiple constraints, calculating valence.

1. Introduction

In addition to describing facts and events, texts often communicate information about the attitude of the writer or various participants towards an event being described. Salient clues about attitude are provided by the lexical choice of the writer but, as discussed below, the organization of the text also contributes critical information for attitude assessment. We start from the current work in this area that concentrates mainly on the negative or positive attitude communicated by individual terms (Edmonds and Hirst, 2002; Hatzivassiloglou and McKeown, 1997; Turney and Littman, 2002; Wiebe et al. 2001). We argue that this approach is incomplete and often gives the wrong results when implemented directly. We describe how the base attitudinal valence of a lexical item can be modified by context and propose a simple "proof of concept" implementation for some contextual shifters.

2. From Simple Valence to Contextually Determined Valence

2.1 Simple Lexical Valence

Examples of lexical items that communicate a negative or positively attitude (*valence*) can be found in all open word classes and as multi-word collocations such as *freedom fighter*. Below we have listed some examples of English words which can be readily characterized as positively or negatively valenced[1].

PART OF SPEECH	Positive Valence	Negative Valence
Verbs	Boost, Embrace, Ensure, Encourage, Delight, Manage, Ease	Conspire, Meddle, Discourage, Fiddle, Fail, Haggle
Nouns	Approval, Benefit, Chance, Approval Benefit, Credit, Favor, Freedom, Hope	Backlash, Backlog, Bankruptcy, Beating, Catastrophe
Adjectives	Attractive, Better, Brave, Bright, Creative, Dynamic,	Annoying, Awry, Arbitrary, Bad, Botched,
Adverbs	Attractively, …	Annoyingly, …

Table 1. Examples of words with non-neutral valence.

2.2 Lexical Valence in Texts

To illustrate how lexical valence influences interpretation, let us look briefly at three short texts.[2] While all of the texts communicate the same denotative information, the connotative force of each version is different. In the first text, the protagonist is an unremarkable young man, in the second text, he is a much friendlier, warmer sort of chap while he emerges in the third text as a juvenile delinquent[3]:

Text 1. *The **eighteen year old** **walked** through the **part of town where he lived**. He **stopped for a while** to **talk** with people on the street and then **went** to a store for some **food** to bring to the **small apartment** where he **lived** with some **people he knew**.*

[1] Not all terms can be characterized along this dimension: many terms are essentially neutral.

[2] Space constraints and the difficulty of finding short texts that exemplify important complex cases while presenting few other distractions oblige us to construct our example.

[3] Notation: Relevant terms are bold; positive terms are marked with a +; negative terms are marked with a -; comparable neutral terms are underlined.

Text 2. *The* ***young man***$^+$ ***strolled***$^+$ *through* ***his neighborhood***$^+$*. He* ***lingered***$^+$ *to* ***chat***$^+$ *with people on the street and then* ***dropped into***$^+$ *a* ***shop***$^+$ *for some* ***goodies***$^+$ *to bring* ***home***$^+$ *to the* ***cozy***$^+$ *place which he* ***shared***$^+$ *with some* ***friends***$^+$*.*

Text 3. *The* ***teenaged male***$^-$ ***strutted***$^-$ *through his* ***turf.*** *He* ***loitered***$^-$ *to* ***shoot the bull***$^-$ *with people on the street and then* ***ducked***$^-$ *into a* ***dive***$^-$ *for some* ***grub***$^-$ *to bring to the* ***cramped hole-in-the-wall***$^-$ *where he* ***crashed***$^-$ *with some* ***cronies***$^-$*.*

The difference in perlocutionary force among these texts emerges solely from the combined effects of the choice of synonyms (or near synonyms) chosen to depict the persons, events and situation involved.

Observations such as these have led researchers to classify terms as positive or negative. The simple computation of the attitude expressed in a text would then consist of counting the negative and positive instances and decide on the basis of the highest number.
To see that the simple counting will not work for many texts, consider the following example (from *The Economist*):

Of course, that would not stop deregulation of the power industry altogether. The ***blunderbuss***$^-$ *of state initiatives will see to that. However, by prolonging* ***uncertainty***$^-$*, it would* ***needlessly***$^-$ ***delay***$^-$ *the arrival of the* ***bonanza***$^+$ *of* ***benefits***$^+$ *that consumers* ***deserve***$^+$*, and give them* ***legitimate***$^+$ *grounds for their* ***cynicism***$^-$*.*

While there are six negative lexical items (marked with -) and only four positive items (marked with +) in this text, readers do not conclude that the author is negative about "deregulation". In fact, the writer views deregulation positively. Clearly, then, the full story of how lexical items reflect attitudes is more complex than simply counting the valences of terms would suggest. In the reminder of this paper, we will propose a number of ways in which the basic valence of individual lexical items may be strengthened or weakened by context provided by (1) the presence of other lexical items, (2) the genre type and discourse structure of the text and (3) cultural factors.

In looking at texts, it is clear that lexical items can be strongly positive or negative or somewhat strong or weak or "hint" at a positive or negative connotation. Therefore, characterizing terms in binary terms as either positive or negative as we have done so far is too crude. Believing that it would be desirable to have a more fine-grained classification, we have adopted a slightly more sensitive scale with three positive and three negative values. In the notation we adopt in this paper, therefore, we assume that words like ***clever*** and ***successful*** are marked +2 in the lexicon. Negatively valenced items are marked -2. It should be kept in mind, however, that this scheme falls far short of an adequate solution to this problem.

3. Contextual Valence Shifters

3.1 Sentence Based Contextual Valence Shifters

While some terms in a text may seem to be inherently positive or negative, we shall show how others change base valence according to context – receiving their perlocutionary force either from the domain of discourse or from other lexical items nearby in the document. In the remainder of this paper we will discuss a number of interacting factors that make the determination of the point of view that an author expresses in a document difficult. We will begin with a survey of several lexical phenomena that can cause the valence of a lexical item to shift from one pole to the other or, less forcefully, to modify the valence towards a more neutral position.

3.1.1 Negatives and Intensifiers

The most obvious shifters are negatives.[4] How *not* can flip the valence of a term has been discussed in the computational literature (Das and Chen, 2001; Pang, Lee and Vaithyanathan 2002). However, in addition to *not*, negatives can belong to various word classes. Simple negatives include *never, none, nobody, nowhere, nothing*, and *neither*. For example:

*John **is** clever* versus *John **is not** clever.*
*John **is** successful at tennis* versus *John **is never** successful at tennis.*
Each of them is successful versus ***None** of them is successful.*

Combining positively valenced words with a negation such as ***not*** flips the positive valence to a negative valence. For example[5]:

clever +2 combined with *not* ⇒ *not clever* -2
successful +2 combined with *not* ⇒ *not successful* -2

The combination of a positive evaluator with a negation turns the evaluation as a whole into a negative one. Inversely the combination of a negative evaluator with negation turns the whole into a positive evaluation (e.g., "*He is **not stupid***").

Not all modifiers switch the valence. Intensifiers such as the *rather* in *rather efficient* and the *deeply* in *deeply suspicious* act to weaken or strengthen the base valence of the term modified. *Rather* weakens the force of a term and *deeply* enhances it (Riloff and Wiebe, 2003). We can calculate their effect by adding or subtracting a 'point' to/from the base valence of a term.

Suspicious `-2` ⇒ *deeply suspicious* `-3`
Efficient `+2` ⇒ *rather efficient* `+1`

As with the negative shifters, intensifiers can belong to all open lexical classes. In addition to adverbs, quantifiers such as *few*, *most*, and nouns such as *lack (of)* also exist.

3.1.2 Modals

Language makes a distinction between events or situations which are asserted to have happened, are happening or will happen (*realis* events) and those which *might, could, should, ought to*, or *possibly* occurred or will occur (*irrealis* events). **Modal operators** set up a context of possibility or necessity and in texts they initiate a context in which valenced terms express an attitude towards entities which do not necessarily reflect the author's attitude towards those entities in an actual situation under discussion. Therefore, in computing an evaluation of the author's attitude, terms in a modal context should not be treated precisely as terms in a realis context.

Assume the realis sentences: *Mary is a terrible person. She is mean to her dogs*. *Terrible* and *mean* are negatively valenced terms. The score for each of the sentences is `-1`. However, the sentence *If Mary were a terrible person, she would be mean to her dogs,* asserts neither that *Mary*

---

[4] Of course for a shift in attitude to take place there has to an attitude expressed in the first place. A simple sentence such as "*John is home*" might express a simple fact without betraying an attitude (i.e. the attitude score is 0). When negated, as in "*John is not home*", there is no shift in attitude (i.e. the negation of 0 is 0).
[5] While it is a simplification to take the scope of a negative as always a whole clause, we will assume this here.

is a terrible person nor *that she is mean to her dogs.* On the contrary, the force of *would* suggests that *she is not mean to her dogs* while the *If* sets up a context in which *Mary is* not necessarily *a terrible person.* In fact, we tend to believe that she is not *terrible* at all. Therefore, the modal operators neutralize the base valence of *terrible* and *mean,* resulting in a re-computed value of 0 for the modal version.

3.1.3 Presuppositional Items

Often words shift the valence of evaluative terms through their presuppositions. This is typical for adverbs like *barely* as shown by comparing "It is sufficient" with "It is *barely sufficient.*" "Sufficient" is a positive term, "barely sufficient" is not: it presupposes that better was expected. These terms can introduce a negative or a positive evaluation even when there are no other evaluative terms around, as in *He got into Foothill College* versus *He barely got into Foothill College* or *He got into Harvard* and *He even got into Harvard.* Words like *barely* and *even* will be marked in the lexicon as evaluation words that interact with other terms. For instance, in the sentence *It was barely sufficient,* the evaluation of the combination is negative. Examples of nouns that act like shifters are *failure* and *neglect*. In the phrase '*failure to succeed*', for example, the force of the meaning of *failure* transforms the positive valence of *succeed* into a negative property. The expression as a whole counts as negative.

The same observations can be made with respect to verbs like *fail, omit, neglect*…. They not only convey the information that something did not happen but also that the author was expecting it to happen and that this not borne out expectation has negative consequences as illustrated by *He stayed around* versus *He failed to leave.*[6]

3.1.4 Irony

Sometimes the contributions made by various lexical items combine in ways that cannot be accounted for in the ways described above. For example, in the ironic sentence *The very brilliant organizer failed to solve the problem*[7]*,* the extremely positive connotation of *very brilliant* is turned against itself by the meaning of the sentence. We account for this phenomena by assuming that in the lexicon *brilliant* will be marked as +2, *very* will increase the base valence of the expression to +3; *fail* will be marked as negative and the expression *solve the problem* will be marked positive. Evaluative terms under the scope of *fail,* such as s*olve the problem* will be marked 0; entities whose existence is not denied by the use of *fail* but to whom failure is ascribed will turn negative. In this case, the base score was 0, however, very brilliant goes from positive to negative and *solve the problem* is neutralized*,* while *fail* remains negative*.* The adjusted score is: -4.

---

6 Often the use of *fail* leads to an indirect negative evaluation of the person to whom the failure is attributed. This can be exploited in irony (see below).

7 Note that when we add *even*, the situation changes again. The sentence is not necessarily ironic. Items under the scope of words like *even* are neutralized. So the sentence *Even the* brilliant *organizer failed to* solve the problem*.* is scored -1 for *fail* only.

brilliant	+2		*Original valence is adjusted by* very
very brilliant	+3	-3	*adjusted because of* fail
failed	-1	-1	
solve the problem	+1	0	*neutralized by* fail
total score:		-4	

Table 2. Valence calculation for "The very brilliant organizer failed to solve the problem."

3.2 Discourse Based Contextual Valence Shifters

3.2.1 Connectors

Connectors such as *although, however, but, on the contrary, notwithstanding* etc. can both introduce information, and act on information elsewhere in the text to mitigate the force of that information[8]. For example, take the sentence *Although Boris is* ***brilliant*** *at math, he is a* ***horrible*** *teacher*. While the statement *Boris is brilliant at math* positively assesses Boris' math skills, the force of *although* combined with the negative assessment in the sentence's main clause *he is a* ***horrible*** *teacher* effectively negates the positive force of the evaluation as applied to Boris. In computing the author's attitude towards Boris, therefore, the effect of *although* is to neutralize the effect of the positive assessment, resulting in a negative assessment score for the sentence. Let's follow that along step-by-step to make the claim clear:

Although Boris is brilliant at math, he is a horrible teacher.

Base valence of terms:		Adjusted computation:	
brilliant	+2	***(Although) brilliant***	0
horrible	-2	***horrible***	-2
total score:	0	total score:	-2

Table 3. Example of valence adjustment based on discourse connective.

In this example we also see how the micro organization of the discourse makes a difference: the positive effect of ***brilliant*** is encapsulated in the embedded clause and does not contribute to the evaluation of the larger unit.

3.2.2 Discourse Structure and Attitude Assessment

A third discourse level valence adjuster included in this paper concerns *discourse structure* itself. There are two basic discourse relations of interest to us here: *lists* and *elaborations*. Some discourse constituents are linked to others in a list in which each constituent encodes a similar relationship to some more general concept and other constituents that give more detailed information of some sort about material encoded in constituents preceding them in the linear organization of the text. These earlier constituents structurally dominate the elaborating constituents (Grosz and Sidner, 1986; Mann and Thompson, 1988; Polanyi and Scha, 1984). Of

[8] As was noticed by Hatzivassiloglou and McKeown (1997), the construction *Adj1 but Adj2* can be used to determine the valency of one adjective if the valency of the other one is known.

interest to us here is how base lexical valence scores are modified by their position in a hierarchical discourse structure.

In an Elaboration, a constituent gives more detail about another constituent which is in a structurally accessible position in a discourse stack. For instance in *John walks a lot. Last month he walked 25 miles on Tuesdays,* the second sentence illustrates the concept expressed in the dominating sentence. When valence information is introduced in a dominating sentence, the elaborations reinforce its effects. For example, lexical valence information is introduced by the use of *terrific* in the dominating sentence in the following passage:

John is a ***terrific***$^{+}$ *athlete. Last week he walked 25 miles on Tuesdays. Wednesdays he walked another 25 miles. Every weekend he hikes at least 50 miles a day.*

Each of the dominated constituents is itself neutrally valenced. However, in this text, each is an example of John's terrific athleticism. Therefore, the positive valence of *terrific* is inherited by each subsequent new example. Effectively, the force for this one instance of the positively valenced term *terrific* as applied to *John* is greatly strengthened when the sentence is treated in its discourse context rather than as an independent expression.

3.2.3 Multi-entity Evaluation

Up to now we have looked at the effects that context can have on the evaluation of one single entity. But in most complex documents a wide variety of entities are discussed – some of which might be evaluated positively and others negatively. For example, a product reviewer discusses one negative aspect of a product extensively in a review which was otherwise very positive about many other features. In this case, it would be incorrect to assume that the reviewer was negative towards the product because of having described one negative feature in some detail. In such a case, simple methods of comparing the number of positive terms versus the number of negative terms could result in a faulty assessment of the reviewer's attitude towards the product. No simple correlation need obtain between the length at which a particular aspect of situation is discussed and the weight that discussion plays in an overall assessment.

3.2.4 Genre and Attitude Assessment

The assessment of author attitude may be complexly related to the genre of the communication in which valence marked terms occur. For example, any use of evaluative language in a document in which such assessments seldom occur will carry more weight than would otherwise be the case. Similarly, the presence of valence carrying items in a text by an author or found in a text type associated with the use of highly evaluative language may carry less weight. As we show below, assessing attitude in a document in which there are various participants "speaking" in a text can be at issue as well.

3.2.5 Reported Speech

Take the sentence *Mary was a slob.* The base valence of this sentence is –1, since ***slob*** is a negatively valenced term. Now, consider, *John said that Mary was a slob.* Here the author asserts that John said something unflattering about Mary, not that the author accepts John's assessment. However, information later in the text could force its inclusion as in *John said that Mary was a slob and he is right.* In this case, the negative valence attached to *slob* will be counted along with the positive valence of *right.* To illustrate consider this text:

The utilities argue that they ***performed well***$^{+}$*. But the public still remembers those* ***miserable***$^{-}$ ***rotten***$^{-}$ *nights.*

Both *argue* and *remembers* are Reported Speech and Thought operators. Therefore, the valence of the reported material is not ascribed to the author (Wiebe, Wilson and Bell, 2001; Wiebe, Wilson, Bruce, Bell and Martin, 2004) but to the utilities and the public respectively. The positive and the negative valences do not cancel each other out. The text is not neutral; it is positive in relation to utilities and and negative in relation to public. We need two different counters one for the utilities and one for the public.

Notice also that the weight of both valences is not equal for the larger unit composed of both sentences. By using *But* the author chooses to give more weight to the second point of view reported as this point in the text. A sensitive weighing scheme could be devised to reflect these complex facts (Riloff and Wiebe, 2003).

3.2.6 Subtopics

Sometimes it is possible to split a longer document into subtopics. The point of view of the author can then be made relative to each subtopic. Take for instance the following artificially constructed short text.

Our yearly overview of the situation in Ubitopia.
The economic situation is more than ***satisfactory***$^{+}$*. The leading indicators show a* ***rosy***$^{+}$ *picture. The manufacturing sector is* ***booming***$^{+}$*. Exports have* ***exceeded***$^{+}$ *the* ***wildest expectations***$^{+}$*.*
When one looks at the human rights picture, one is struck by the increase in ***arbitrary***$^{-}$ *arrests, by needless*$^{-}$ ***persecution***$^{-}$ *of* ***helpless***$^{-}$ *citizens and increase of police* ***brutality***$^{-}$*.*

In a text like this, one could link the positive and the negative attitudes to the two subtopics, the economy and the human rights situation. In most cases, this will not be as easy as is the case here, and, even if the text can be clearly divided into subtopics, it is not necessarily the case that all subtopics contribute equally to the overall impression that a text makes. One factor that will influence their contribution is genre, which we discuss next.

3.2.7 Genre Constraints

Movie reviews have been a focus of attention in the document classification community for some time. These texts are known to be notoriously difficult to work with using existing techniques (Pang, Lee and Vaithyanathan, 2002). Problems arise because these texts are composed of two types of information: information about the events and situations in the story and information about the film which has been created to tell the story. Since the question one is interested in primarily interested in having answered by a movie review is *Is this a good movie?* and since the review is prepared by the reviewer to answer this question, it is necessary to separate the description of the entities pertaining to the story from the description of the entities pertaining to the production. Only the valence scores of the entities pertaining to the production should be considered in ascertaining if the review is positive or negative.

Reviews of films loosely follow a set of genre conventions that can be mined for factors which can influence basic valence assignment. For example, movie reviews are often constructed as a quasi-interaction between author and reader. Comments in or about the first or second person reflect information about the film since neither reader nor author are characters in the film. Positional information can also be important: comments at the very beginning or very end of a review are accorded more weight than remarks in less prominent positions.

Let's consider an artificially constructed example based on an excerpt from a movie review taken from MRDb website used in Pang, Lee and Vaithyanathan (2002)[9]:

This film should be ***brilliant***$^+$*. The characters are* ***appealing***$^+$*. ... Stallone plays a* ***happy***$^+$*,* ***wonderful***$^+$ *man. His* ***sweet***$^+$ *wife is* ***beautiful***$^+$ *and* ***adores***$^+$ *him. He has a* ***fascinating***$^+$ ***gift***$^+$ *for* ***living life fully***$^+$*. It sounds like a* ***great***$^+$ *plot, however, the film is a* ***failure***$^-$

	Adjusted Score	
This film should be ***brilliant***$^+$**.**	0	*brilliant* within scope of *should* is 0
The characters are ***appealing***$^+$.	0	*appealing* elaboration under *should* 0
Stallone plays a ***happy***$^+$***, wonderful***$^+$ *man. His* ***sweet.*** *wife is* ***beautiful***$^+$ *and* ***adore****s him. He has a* ***fascinating***$^+$ ***gift***$^+$ *for* ***living life fully***$^+$.	0	*Happy, wonderful, sweet* etc. all refer to storyworld entities and thus are not counted.
It sounds like a ***great***$^+$ *story,*	-1	*However* reverses + valence of *great*
however, the film is a ***failure***$^-$	-1	*failure* is –1
Total Score:	-2	

Table 4. Valence calculation for the movie review.

The adjusted score is –2. The review is negative.

In some cases, we should be able to exploit genre constraints in determining the attitude of authors towards the entities created in the documents. But to do this computationally, the structure that genres impose on documents needs to be determined automatically. This is not yet possible.

4. Conclusion

We have shown that even when the author attention is restricted to one topic/entity/fact, lexical items in a discourse context will interact with one another. An author's attitude cannot be calculated based on individual items. We proposed a calculation of local interactions that improves upon the results of current approaches based on simple counts. We also argued that valence calculation is critically affected by discourse structure. In addition, we discussed cases in which a document describes more than one entity/topic/fact. We showed that, in these cases, the calculation of point of view must be done with respect to each entity separately and must take into account higher order factors such as genre that influence document structure

Taken together, these considerations argue strongly that calculating author attitude must be based on a finer grained analysis of the text on all levels than has been previously proposed.

[9] The authors explain that their context insensitive evaluative lexical methods fail on texts in which the author sets up a deliberate contrast to an expected position. They cannot deal with the mismatch between the base valence of the term and the author's usage.

5. Bibliography

Das, S. and Chen, M. (2001) *Yahoo! For Amazon: extracting market sentiment from stock message boards*. Paper presented in the *8th Asia Pacific Finance Association Annual Conference*, Bankok, Thailand.

Edmonds, P. and Hirst, G. (2002) *Near-synonymy and lexical choice.* Computational Linguistics, 28 (2), 105–144.

Grosz, B. and Sidner, C. (1986) *Attention, Intention, and the Structure of Discourse.* Computational Linguistics, 12 (3), 175–204.

Hatzivassiloglou, V. and McKeown K. (1997) *Predicting the Semantic Orientation of Adjectives.* Computational Linguistics, 174–181.

Hatzivassiloglou, V. and Wiebe, J. (2000) Effects of Adjective Orientation and Gradability on Sentence Subjectivity. Paper presented at Coling 2000. Saarbrucken.

Mann, W. and Thompson, S. (1988) *Rhetorical Structure Theory: Toward a functional theory of text organization.* Text 8 (3), 243–281.

Pang, B. Lee, L. and Vaithyanathan, V. (2002) *Thumbs up? Sentiment Classification using Machine Learning Techniques*. In *Proceedings of the 2002 Conference on Empirical Methods on Natural Language Processing.* 79–86.

Polanyi, L. and Scha, R. (1984) *A Syntactic Approach to Discourse Semantics.* Paper presented at *6th International Conference on Computational Linguistics*, Stanford, CA.

Riloff, E. and Wiebe, J. (2003) *Learning Extraction Patterns for Subjective Expressions.* In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. 105–112.

Tong, R. (2001) *An operational system for detecting and tracking opinions in on-line discussion.* Paper presented at the *SIGIR Workshop on Operational Text Classification*.

Turney, P. and Littman, M. (2002) *Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus.* Technical Report, National Research Council Canada, Institute for Information Technology, ERB-1094, NRC-44929.

Wiebe, J., Wilson, T., and Bell, M. (2001) *Identifying Collocations for Recognizing Opinions.* Paper presented at the *ACL/EACL '01 Workshop on Collocation*, Toulouse, France, July 2001.

Wiebe, J., Wilson, T., Bruce, R., Bell, M., and Martin, M. (2004) *Learning subjective language.* Computational Linguistics 30 (3).

Chapter 2

Conveying Attitude with Reported Speech

Sabine Bergler
Concordia University
Department of Computer Science and Software Engineering
1455 de Maisonneuve Blvd. West
Montreal, Quebec, H3G 1M8, Canada
Email: bergler@encs.concordia.ca

Abstract

Attribution is a phenomenon of great interest and a principled treatment is important beyond the realm of newspaper articles. The way natural language has evolved to reflect our understanding of attribution in the form of reported speech can guide investigations into principled representations forming the basis for shallow text mining as well as belief revision or maintenance.

Keywords: attribution, reported speech, reliability of information, argumentative structure, profile structure, potential belief space.

1. Introduction

Society has developed a multitude of mechanisms that serve to authenticate items, and in particular information. Signatures authenticate letters, paintings, and seal contracts. Imprints on money, seals, and forms make them official. Insignia establish membership in certain groups, as do uniforms and religious symbols. Information, likewise, has well established mechanisms of authentication, which vary slightly from society to society. The Native American language Pawnee has four different prefixes that obligatorily have to mark statements for their reliability (hearsay, reasonably certain but not witnessed directly, leaving room for doubt, or mere inference) (Mithun, 1999). And while the number and type of such evidential markers differ in different languages, hearsay is maybe the most widespread one.

English and most European languages do not have a system of evidential morphology, but mark hearsay and other evidentiality at the syntactic level. Reported speech, both in form of direct quotation (*... and then she said "I have to go."*) or indirect paraphrases (*... and then she said that she had to go.*), is the most formalized register. Reported speech is most prominent in newspaper articles, where it can occur in up to 90% of the sentences of an article. Computational linguistic treatments of newspaper articles usually ignore reported speech, either by omitting the material entirely, or by ignoring its evidential status. This paper argues that reported speech segments information into coherent subunits, called *profiles* after (Bergler, 1992). Different profiles can imply different credibility of the source of the information, different roles of the source in an argumentative structure, or a different context (temporal or other). An extended example illustrates profiling on a product review article. This paper concludes that the mechanism of profiling (and its proper analysis) should be extended beyond reported speech to all explicit attributions, such as newsgroup messages, etc.

2. Evidential Analysis of Reported Speech

Reported speech is characterized by its syntactic structure: a matrix clause, containing at least the source as subject NP and a reporting verb, embeds the information conveyed in a complement clause. The complement is optionally introduced by "that" for indirect reported speech, and it is surrounded by quotation marks for direct reported speech. As argued in (Bergler, 1992), the complement usually conveys the *primary information* in newspaper articles and most other genres. In fact, the case where the matrix clause bears the major information, namely that something had been uttered by somebody under certain circumstances without the utterance itself being of importance, is rare (but see (Clark and Gerrig, 1990) for examples). The syntactic dominance of the matrix clause shows the semantic importance of the contained *circumstantial information* (Bergler, 1992), the who, when, where, and how. But the natural propositional encoding of the complement clause as embedded in the matrix clause is not suitable. Rather, the information of the matrix clause should be seen as a meta-annotation for interpreting the primary information differently in different contexts and for different purposes. Thus the a priori trust a Republican reader has in utterances by Cheney is different from a Democrat's. And a text will be interpreted differently at the time of the events unfolding and after additional information is known. This variability of the pragmatic force of the matrix clause also suggests that it cannot be "resolved" at the time of first text analysis, but has to remain attached in a form close to the original for further analysis. (Bergler, 1992) gives a general linguistic treatment of reported speech. This paper presents, in contrast, one particular implementation of the general representation for further automatic analysis. The underlying assumption is that the further processing will be by an information extraction or mining system that works with shallow, possibly statistical techniques. But the representation does not preclude the deeper linguistic analysis outlined in (Bergler, 1992).

Politics & Policy: Democrats Plan Tactic to Block Tax-Cut Vote
Threat of Senate Filibuster Could Indefinitely Delay Capital-Gains Package

(S_1) Democratic leaders have bottled up President Bush's capital-gains tax cut in the Senate and may be able to prevent a vote on the issue indefinitely.

(S_2) Yesterday, Sen. Packwood acknowledged, "We don't have the votes for cloture today."

(S_3) The Republicans contend that they can garner a majority in the 100-member Senate for a capital-gains tax cut.

(S_4) They accuse the Democrats of unfairly using Senate rules to erect a 60-vote hurdle.

(S_5) Democrats asserted that the proposal, which also would create a new type of individual retirement account, was fraught with budget gimmickry that would lose billions of dollars in the long run.

Figure 1. Adapted from Jeffrey H. Birnbaum, The Wall Street Journal, 10/27/89.

As Figure 1 demonstrates, the role of reported speech is attribution: the statement does not assert as `true' what amounts to the information content of the sentence, but a situation in which this content was proffered by some source. This device can be used both to bolster a claim made in the text already, and to distance the author from the attributed material, implicitly lowering its credibility (Anick and Bergler, 1992). Thus the credibility or reliability of the attributed information is always in question for reported speech and other attributions. If the attribution is used to bolster a claim already made by citing a particularly strong source for endorsement, ignoring the fact that an explicit attribution was made will do no great harm. This is in fact a frequent case in the type of newspaper articles typically used for large-scale system development and testing (as in MUC, TREC, DUC, etc.) and this is why ignoring attribution has been tolerable. But when a text is argumentative (opposing two or more points of view on a topic), speculative (when the final outcome of an event is not yet decided and the text uses different sources as predictors), or presents a personal opinion or experience, text meaning depends on proper attribution recognition (Bergler, 1995a). Argumentative or speculative text structure is not limited to newspaper articles. Scientific articles, too, use reported speech for this purpose, but in a different rhetorical style. And multi-participant political analysis segments on newscasts form the same phenomenon: different opinions are identified with different individuals and contrasted, even though we might term it broadcast speech, rather than reported speech. Interestingly, broadcast speech retains the required elements of reported speech, in that it is always anchored by the identity of the source and the circumstances of the utterance (date, occasion, place, etc.) as they are relevant to its analysis. The reported material is always literal and quoted, of course, but has still undergone an editing process, extracting the broadcast speech from a larger interview and potentially juxtaposing material that the source did not intend to. Thus the simple fact that no paraphrasing is involved does not make broadcast speech necessarily truer to the original than reported speech.

Reported speech in newspaper articles can be detected and analyzed without a complete syntactic analysis, using shallow means and standard tools. In a feasibility study, Doandes (2003) presents a knowledge-poor system that identifies sentences that contain reported speech in Wall Street Journal texts and analyzes them into structures inspired by (Bergler, 1992) and illustrated in Figure 2.

The system works in a shallow environment: Built on top of ERS (Witte and Bergler, 2003), it has access to slightly modified versions of the Annie tools distributed with GATE (Cunningham, 2002) and an in-house NP chunker and coreference resolution module. The NP chunker relies on the Hepple tagger (Hepple, 2000) and Annie Gazetteer, the coreference module has access to WordNet (Fellbaum, 1998).

`BP`	basic profile
`OTHERCIRC`	circumstantial information other than source and reporting verb
`PARAPHRASE`	paraphrased material, usually complement clause in case of indirect reported speech
`REPSOURCE`	source, in active voice the matrix clause subject
`REPORTEDSPEECH`	complement clause
`REPVERB`	reporting verb, main verb in matrix clause
`QUOTEDSPEECH`	material in quotation marks

Figure 2. Template for representing reported speech sentences in (Doandes, 2003).

Doandes uses part-of-speech tags to identify main verb candidates. In a detailed analysis of verb clusters, she determines main verbs and compares them against a list of likely reported speech verbs. In case a reported speech verb is found, the sentence pattern (with complete part-of-speech annotations, annotations for NPs, and annotations for verb groups) is compared to the possible patterns for reported speech constructions as described in (Quirk et al., 1985). Figure 3 gives the resulting *basic profile* for the sentence: *Yesterday, Sen. Packwood acknowledged, "We don't have the votes for cloture today."*

```
BP
OTHERCIRC Yesterday,
PARAPHRASE
REPSOURCE Sen. Packwood
REPORTEDSPEECH, "We don't have the votes for cloture today."
REPVERB acknowledged
QUOTEDSPEECH "We don't have the votes for cloture today."
```

Figure 3. Example representation in (Doandes, 2003).

The development corpus consisted of 65,739 sentences from the Wall Street Journal, the test corpus of 2,404 sentences taken mainly from the Wall Street Journal, with a few articles from the DUC 2003 corpus of newspaper articles (DUC, 2003). 513 occurrences of reported speech were found and precision is 98.65%, recall is 63%. The analysis into basic profiles incurred some mistakes (such as retaining only part of the subject NP in the SOURCE slot). Using a strict notion of correctness for the entire basic profile, the performance drops to 87% precision and 55% recall.

Many recall problems are linked to limitations of the particular implementation, such as tagging errors, the NP chunking process (the NP chunker splits heavy NPs into several smaller chunks, thus occasionally obfuscating the reported speech pattern), and an incomplete list of reported speech verbs. (Doandes, 2003) works from a simple list of candidate reported speech verbs with no attempt at word sense disambiguation. The results seem satisfactory for the evaluation corpus, but will not necessarily hold outside the newspaper genre. (Wiebe et al., 1997) report on the difficulty of distinguishing *private state, direct speech, mixed direct and indirect speech, other speech event, other state or event*. Most of these categories describe attributions and thus do not need to be distinguished for profile structure at the level described here, even though their distinction would refine the use of the profile for subsequent processing.

3. Profile Structure

Figure 1 is typical for newspaper articles: information from two different points of view, here Democrats and Republicans, is interleaved. Ideally, an automatic system would group S_1 and S_5 into one profile, and S_2, S_3, and S_4 into another, effectively grouping Democrats versus Republicans. This is, however, not possible with shallow techniques. S_1 is not a reported speech sentence and thus does not generate a profile. World knowledge is required to infer that *Sen. Packwood* speaks for the *Republicans* in this article, but pronoun resolution techniques allow *they* to be resolved to *Republicans*, creating a merged profile from S_3 and S_4, enabling interpretation of *a 60-vote hurdle* in the context of S_3.

Profile structure is complementary to both rhetorical structure (cf. Marcu, 1997) and text structure (cf. Polanyi et al., 2004). It creates another type of context, which is coherent with respect to underlying processing assumptions, such as reliability of the source, or, as seen above, inferential assumptions (*60-vote hurdle*). For a more detailed discussion, see (Bergler, 1995a). The profile structure for the text in Figure 1 is given in Figure 4.