Adjectives and Adverbs as Indicators of Affective Language for Automatic Genre Detection

Robert Rittman and **Nina Wacholder**

School of Communication, Information and Library Studies
Rutgers, The State University of New Jersey
New Brunswick, New Jersey

Abstract. We report the results of a systematic study of the feasibility of automatically classifying documents by genre using adjectives and adverbs as indicators of affective language. In addition to the class of adjectives and adverbs, we focus on two specific subsets of adjectives and adverbs: (1) trait adjectives, used by psychologists to assess human personality traits, and (2) speaker-oriented adverbs, studied by linguists as markers of narrator attitude. We report the results of our machine learning experiments using Accuracy Gain, a measure more rigorous than the standard measure of Accuracy. We find that it is possible to classify documents automatically by genre using only these subsets of adjectives and adverbs as discriminating features. In many cases results are superior to using the count of (a) nouns, verbs, or punctuation, or (b) adjectives and adverbs in general. In addition, we find that relatively few speaker-oriented adverbs are needed in the discriminant models. We conclude that at least in these two cases, the psychological and linguistic literature leads to identification of features that are quite useful for genre detection and for other applications in which identification of style and other non-topical characteristics of documents is important.

1 INTRODUCTION

This paper reports on the use of adjectives and adverbs to discriminate text genres characterized by affective expressions (e.g., fiction) from genres in which affective expressions are typically inappropriate (e.g., academic writing).[1] We adopt the definition of genre given by Lee [2].

> [G]enre is a document-level category assigned on the basis of **external** criteria such as intended audience, purpose, and activity type, that is, it refers to a conventional, culturally recognised grouping of texts based on properties other than lexical or grammatical (co-)occurrence features, which are, instead, the **internal** (linguistic) criteria forming the basis of text type categories.

Thus, a news report is intended to inform, an editorial or opinion piece is intended to persuade, and a novel is intended to entertain.

The paper is organized as follows. First, we review discriminating features selected in automatic genre classification research. In Section 3, we summarize how adjectives and adverbs are generally indicative of affective language, and describe the characteristics of two small subsets of adjectives (trait adjectives) and adverbs (speaker-oriented adverbs). In Section 4, we describe our methodology for discriminating documents by genre using these features. In Section 5, we present our results. In Section 6, we discuss our conclusions and provide direction for future work.

2 FEATURE SELECTION IN GENRE CLASSIFICATION RESEARCH

In previous research in genre discrimination, researchers have focused on identifying any features that are useful in discriminating genres. Toward this end, they have identified discriminating features of four basic types: (a) syntactic (parts of speech, e.g., adverbs, nouns, verbs, and prepositions), (b) lexical (terms of address, e.g., *Mr*., *Mrs*., *Ms*.; content words; most frequent words in a corpus, e.g., *the*, *of*, *and, a*); (c) character-level (e.g., punctuation, character count, sentence count, word length in characters); and (d) derivative (ratio measures, e.g., average words per sentence, average characters per word, type/token ratio).

They have applied these features to discriminate different sets of documents and different genres. Using a set of features that were relatively easy to identify automatically in combination with a machine learning method, and working with 500 documents from the Brown Corpus, Karlgren and Cutting [3] selected a set of 20 features, such as first person pronouns, adverbs, prepositions, and nouns; characters per document; average words per sentence; and type/token ratio. Similarly, Kessler at al. [4] classified 500 documents from the Brown Corpus with 55 features (lexical, character-level, and derivative features). Using 500 documents from the LIMAS German corpus, Wolters and Kirsten [5] took a hybrid approach, combining the traditional IR "bag of words" method with a natural language processing method they called, "bag of ‘tagged’ words." They represented documents as vectors of the frequency of content word lemmas and function words, and combined it with part of speech information. Inspired by research in author attribution, Stamatatos et al. [6] selected the 50 most common words in the British National Corpus (e.g., *the*, *of*, *a*, and *in*), as well as eight of the most frequent punctuation symbols (period, comma, colon, semicolon, quotes, parenthesis, question mark, and hyphen). Using a subset of these features Ng et al. [7] selected four punctuation

[1] These results are taken from a much larger study by Rittman [1] of automated classification of documents by genre using adjectives and adverbs as discriminating features.

marks (comma, period, colon, and semicolon) to classify a collection of *Wall Street Journal* and *Federal Register* documents in an investigation of features independent of syntax and semantics. Subject-classified and genre-classified training data was used by Lee and Myaeng [8] to select features based on three criteria: (a) find terms that occur in many documents belonging to one genre which are distributed evenly among all subject classes; (b) eliminate terms that are specific to a particular subject; and (c) downgrade terms that are common to many genres. Web-based technology features based on HTML tags and URL information were selected by Lim et al. [9] in addition to features used by other researchers (e.g., part of speech, punctuation, average words per phrase, and frequency of content words). More than one-hundred features (including syntactic, lexical, and character-level features) were selected by Santini et al. [10] and Santini [11] to address the problem of emerging genres in the Web. The problem of Web genre detection was also addressed by zu Eissen and Stein [12] using thirty-five derivative features, such as average number of mail links, average number of help symbols, and average number of various parts of speech (e.g., nouns, verbs, prepositions, and adverbs). Finally, Finn and Kushmerick [13] classified two sets of Web-generated corpora representing documents as (a) a bag-of-words (vector indicating the presence or absence of a word), (b) part-of-speech statistics (vector of 36 parts of speech features); and (c) text statistics (e.g., average sentence length, average word length, and frequency of punctuation).

The approach to genre identification that characterizes these studies might be called a 'bag of features' approach: researchers applied machine learning techniques to any features that could be identified automatically. Since the focus of their research was genre identification, this approach was completely appropriate. But the use of bags of features, along with different sets of documents and genres, has made it difficult to systematically study the contribution of affective language to genre identification. Only one of the studies described above (Wolters and Kirsten [5]) specifically mentioned that adjectives and adverbs are useful for distinguishing genre, as opposed to a mix of many kinds of features they tested. Furthermore, the authors report their results using the standard measure of Accuracy; this measure does not take into consideration the impact of the percentage of documents that belong to each class on the outcome; this too makes it hard to compare results.

In what follows, we discuss the characteristics of adjectives and adverbs that make them particularly useful for identifying expressions of affect and assess their contribution to automatic genre classification.

3 ADJECTIVES AND ADVERBS AS FEATURES OF AFFECTIVE LANGUAGE

As a grammatical category, adjectives modulate the meaning of nouns by emphasizing important or surprising properties of the noun being modified (e.g., a *safe* / *historical* / *unusual* building). The properties that are highlighted frequently represent a judgment or opinion. The statement, *She wrote a poem*, is a statement of (presumed) fact. The statement, *She wrote a beautiful* / *horrendous* poem, mixes a statement of fact with human judgment. Research indicates a correlation between human perceptions of subjectivity and the occurrence of adjectives in (a) sentences (Bruce and Wiebe [14], Wiebe [15], and Wiebe et al. [16]) and (b) documents (Rittman et al. [17]). This relationship is expected because of the nature of adjectives themselves. Subjective expressions necessarily involve judgments and opinions about people and things, and we frequently use adjectives to express our judgments.

In a similar way, adverbs modulate the meaning of verbs, adjectives, other adverbs, and noun phrases. This is especially true of the many adverbs derived from adjectives by adding the suffix *–ly* (*beautiful* => *beautifully*; *horrendous* => *horrendously*); adverbs typically inherit the subjective connotation of the adjectives from which they have been derived.

Within the larger set of adjectives or adverbs in the context of a sentence, researchers in psychology and linguistics have each indicated a subset of words that appear to be particularly expressive of affect. Psychologists have identified trait adjectives and linguists have identified speaker-oriented adverbs.

3.1 Trait Adjectives

The significance of adjectives in description and judgment has long been noted in psychology. Psychologists use trait adjectives to describe human personality traits (e.g., *nervous, energetic, accommodating*, and *careful*). Trait adjectives are classified by the type of personality they indicate, based on theories of psychology. Using factor analysis on various lists of adjectives Goldberg [18] proposed five dimensions of personality that are generally accepted as the "Big Five": I. Extraversion (*active, assertive, bold*), II. Agreeableness (*agreeable, considerate, cooperative*), III. Conscientiousness (*careful, conscientious, efficient*), IV. Emotional Stability (*imperturbable, relaxed, undemanding*), and V. Intellect (*artistic, bright, complex*). Some researches (e.g., Nowson et al. [19], Argamon et al. [20], and Mairesse et al. [21] have studied the relationship between personality traits of experimental subjects and their use of language features in different genres.

We turn the Big Five on its side and select adjectives that are used by psychologists as indicators of personality as features for genre detection. Although psychologists use these adjectives to scientifically characterize human personality in the context of written and spoken text, when these adjectives are used in non-scientific language, they represent expressions of judgment. What is virtuous to one person may be sinful to another. Furthermore, trait adjectives frequently have an affective connotation; for example, the adjectives *perfidious* and *vulgar* almost always represent a negative judgment while the adjectives *loyal* and *intriguing* almost always represent a positive one. These connotations are pertinent whether the adjectives are used to describe people or some other kind of entity. The trait adjectives in Appendix A (a subset of 44 trait adjectives which we derive from the full list reported by Peabody and De Raad [22], along with the adverbs derived from them (Appendix B), are therefore particularly likely to express affect.

3.2 Speaker Oriented Adverbs

Adverbs that express sentiment typically serve three grammatical functions: disjuncts, adjuncts and subjuncts (Quirk et al. [23]). Disjuncts (1.3) are peripheral to the sentence and "express an evaluation of what is being said either with respect to the form of the communication or to its meaning … [And they express] the speaker's authority for, or comment on, the accompanying clause" (Quirk et al. [23]) For example, in (1.3), *frankly* is a description of the speaker's attitude about the statement, *I am tired*. In contrast, adjuncts (1.1) and subjuncts (1.2) are integrated within the structure of the clause. For example, in (1.1) and (1.2), *slowly* and *kindly* focus internally on the grammatical subject or verb phrase (i.e., they walked *slowly*, you wait *kindly*). This is quite different than the disjunctive use of *frankly*, which focuses externally on the speaker's behavior. As Mittwoch [24] explains, disjuncts are a way to "refer to one's own words." For this reason, disjuncts are referred to as *speaker-oriented adverbs* (SOAs).

(1.1) *Slowly* they walked back home. (Adjunct)
(1.2) Would you *kindly* wait for me? (Subjunct)
(1.3) *Frankly,* I'm tired. (Disjunct)

The potential usefulness of SOAs in identifying expressions of affect is supported by Jackendoff [25] and Ernst [26], who indicate that (a) adverbs can refer to the speaker (narrator), the grammatical subject, or the manner in which an event occurs, (b) sentence position of adverbs affects meaning, (c) adverbs can occur in some positions and not in others, and that (d) adverb phrases can frequently be paraphrased using corresponding adjective phrases. SOAs refer to the speaker of the sentence, subject-oriented adverbs refer to the grammatical subject of the sentence, and manner adverbs refer to the main verb of the sentence. In summary, SOAs provide a grammatical mechanism by which a speaker can insert an indication of mood or attitude at the periphery of the sentence. We use a set of 30 SOAs derived from Ernst [26] (Appendix C).

3.3 Adjectives and Adverbs in Relation to Genre Identification

Since adjectives and adverbs frequently perform some degree of evaluation, it follows that the occurrence of a relatively high number of adjectives and adverbs should indicate the presence of expressions of judgment in a document. This characteristic makes the frequency of adjectives and adverbs in text a likely feature for discriminating genres that include expressions of sentiment and judgment.

Trait adjectives and the adverbs inherited from them frequently have evaluative connotations, at least in context; we expect that they will be most frequent in genres that describe people's behavior, such as fiction. SOAs characterize the narrator's perspective, and are indicative of intent and behavior.

Since genre is indicative of the author's purpose, intended audience, and type of activity (Lee [2]), we explore the contribution of adjectives and adverbs in general, and trait adjectives and SOAs in particular, to the identification of genre.

4 METHODOLOGY

In the first part of this section, we describe the materials used in our study; these include the collection of documents, the genre labels and the features, and the classification problems that we used machine learning methods to solve. In the second part, we describe Accuracy Gain, a measure of the contribution of features to a classification task that is more rigorous than the standard measure of Accuracy used in most genre identification tasks.

4.1 Experimental Materials

To systematically study the relationship of adjectives and adverbs to genre, we needed a set of documents that had been classified by genre and tagged by part-of-speech. Fortunately, the freely available British National Corpus, World Edition (BNC2 [27]) satisfied these requirements. Lee [28, 29] originally assigned each of the 4,054 [2] documents in BNC2 to one of 70 genres; 46 were written and 24 were spoken. However, the large number of genres meant that relatively few documents were assigned to each genre. Davies [30] therefore organized these 70 genres into six supergenres which he labeled *academic*, *fiction*, *news*, *non-fiction*, *other*, and *spoken*. Our experimental task is to assess the contribution of adjectives and adverbs to automatic classification of these six genres.

We consider two basic kinds of genre classification problems. The easier problem is one-against-one; the harder problem is one-against-many. One-against-one discriminates one genre from another (e.g., academic vs. fiction or fiction vs. news). One-against-many discriminates one genre from all other genres in a corpus (e.g., academic vs. fiction, news, non-fiction, other, and spoken, or news vs. academic, fiction, non-fiction, other, and spoken). One-against-one is easier because it considers one document subset of a corpus against another subset, and because only two genre categories are considered. One-against-many is harder because it treats one genre against the entire corpus; the latter consists of documents from multiple genres.

For one-against-one, we chose three of Davies [30] supergenres that are mutually exclusive: academic, fiction and news. This presents three one-against-one classification problems: academic vs. fiction; academic vs. news; and news vs. fiction. For one-against-many, we used the same three supergenres, comparing the performance of classifiers on distinguishing the supergenres from documents consisting of all of the other genres in BNC2. Our one-against-many problems are (1) academic vs. not-academic (fiction, news, non-fiction, other, and spoken), (2) fiction vs. not-fiction (academic, news, non-fiction, other, spoken); and (3) news vs. not-news (academic, fiction, non-fiction, other, and spoken).

We chose the supergenres of academic, fiction, and news for several reasons. First, they are based on Lee's [2] criteria for genre (intended audience, purpose, and activity type). Second, Davies' [30] organization is exclusive. For instance, the set of documents labeled *fiction* excludes academic, news, non-fiction, spoken and other (although non-fiction can be similar in content to other classes, such as academic for instance except it is intended for a different audience). What he calls *spoken* includes everything spoken, regardless of

[2] BNC2 includes 4,054 documents. We exclude one document because it is a duplicate (see Lee [2]).

whether it is academic, fiction, news, non-fiction, or other. We do not use *other* because it is not a conceptually distinct class. Third, selecting only three supergenres directly (academic, fiction, and news) limits the number of discriminant tests to six problems (three one-against-one and three one-against-many), as opposed to 21 problems if we selected all six supergenres (15 one-against-one and six one-against-many). Finally, the supergenres of non-fiction, other, and spoken are treated indirectly in our one-against-many classification problems.

From the complete corpus (*N*=4,053), we randomly selected 50% of the documents for training (*n*=2,029) and 50% for testing (*n*=2,024). Based on this selection, we broke out six document sets. Table 1 shows the number of documents in each set.

Problem	**Doc Sets**	**50% Training**	**50% Testing**	**Total**
		2,029	**2,024**	**4,053**
Acad vs. Fiction	Acad	276	229	505
	Fict	218	246	464
	Total	494	475	969
Acad vs. News	Acad	276	229	505
	News	249	269	518
	Total	525	498	1,023
Fict vs. News	Fict	218	246	464
	News	249	269	518
	Total	467	515	982
Acad vs. NotAcad	Acad	276	229	505
	NotAcad	1,753	1,795	3,548
	Total	2,029	2,024	4,053
Fict vs. NotFict	Fict	218	246	464
	NotFict	1,811	1,778	3,589
	Total	2,029	2,024	4,053
News vs. NotNews	News	249	269	518
	NotNews	1,780	1,755	3,535
	Total	2,029	2,024	4,053

Table 1: Training-Testing Document Sets

In the larger study, Rittman [1] systematically tested the impact of a variety of adjective and adverb features on genre identification on these six problems. First, features were represented as types and tokens. Type indicates whether a word occurs in a document at least once; token indicates the frequency of a word in a document. These two measures give rise to two related representations of features: count and vector. Count is the aggregate of all members of a class in a document. For example, suppose that word_1, word_2 and word_3 all belong to the target class under discussion. And suppose that word_1 occurs three times in a document, word_2 does not occur, and word_3 occurs 7 times. The count of types is 2 (1+0+1). The count of tokens is 10 (3+0+7). The vector approach also identifies types and tokens but represents them a different way. For example, if a word occurs in the document, then type=1; else type=0. If type=1, then token=frequency of the word in the document; else token=0. Thus, in the example above for the three hypothetical words, the vector for the document is (1, 3, 0, 0, 1, 7).

In some cases, the sentence position of a word was marked, such as *sentence-initial position* and *not sentence-initial position* which conceptually is the union of *any sentence position*. Finally, each feature was represented as a *normalized* and a *non-normalized* variable. The normalized variable is calculated by dividing frequency by the total count of words in a document.

We call the various ways of representing features a *method*. Each method includes a unique set of features (such as speaker-oriented adverbs) and different ways of representing the features (such as sentence position, vector or count, or as normalized or non-normalized). The intersection of a method and a classification problem represents a model. Since there are 54 methods for six problems, we analyzed a total of 324 models. In what follows, we report only on a portion of the 324 models identified in the larger study, using only the results of the normalized variables for the adjective and adverb features that are most interesting. Table 2 lists these 17 sets of features

Feature / Method	**Word Class**	**Types in BNC2 ***	**Vector or Count**
SOA1	30 Speaker-Oriented Adverbs (Appendix C) in any sentence position using a vector length of 60	30	V
SOA2	SOA1 in sentence-initial+not-sentence-initial position using a vector length of 120	30	V
SOA3	Count of SOA1 in any sentence position	30	C
RB	All words tagged in BNC2 as adverbs	9,324	C
RB-ly	All RB ending in *ly*	6,372	C
JJ	All words tagged in BNC2 as adjectives	147,038	C
JJ1	Subjective adjectives identified by [31] (in BNC2)	1,322	C
JJ2	Trait Adjectives full list derived from [22] (in BNC2)	732	C
JJ3	Subset of JJ2 (Appendix A) identified in pilot study as indicative of subjectivity	44	V / C
RB1	Trait Adverbs derived from JJ2 by adding *ly*	539	C
RB2	Subset of RB1 (and JJ3) (Appendix B) using vector length of 72, also as a count	36	V / C
JJ3+RB2	Union of JJ3+RB2 using a vector length of 160	80	V

SOA1+JJ3 +RB2	Union of SOA1+JJ3+RB2 (Appendix A, B, C) using a vector length of 220	110	V
All-JJ&RB	SOA3, JJ, JJ1, JJ2, JJ3, RB, RB-ly, RB1, RB2	165,437	C
NN	Nouns	430,415	C
VB	Verbs	60,083	C
Punc	Punctuation	71	C
* For each feature we consider both types and tokens. Thus the number of variables for each feature is doubled.			

Table 2: Selected Features

We include the count of all adjectives and adverbs that were tagged as such in BNC2, including all adverbs ending in *–ly*, and all adjectives identified by Wiebe [31] as *subjective adjectives*. For our vector experiments, we select the 30 speaker-oriented adverbs derived from [26] (Appendix C), the subset of 44 trait adjectives derived from Peabody and De Raad [22] (Appendix A), and a list of 36 *trait adverbs* derived from Appendix A by adding *-ly* (Appendix B). We also combine these three lists of words in a union set of 110 adjectives and adverbs (Appendix A, B, and C). We also included nouns, verbs and punctuation as a benchmark to compare the performance of models using our adjective and adverb features.

Each set of features in Table 2 was used to build different models using tools for discriminant analysis provided by SPSS (version 11.0). The machine learning method of discriminant analysis is a widely used classification method of multivariate statistics used in genre classification work by Karlgren and Cutting [3], Stamatatos et al. [6], and Ng et al. [7]. For each classification problem, we test the performance of various sets of features and methods.

4.2 Accuracy Gain [3]

The standard measure of performance for classification problems is Accuracy [32]. Simply put, Accuracy is the fraction of correctly predicted cases. However, Accuracy does not consider the proportion of members of a particular class. As such, it does not take into account the most rigorous and expected baseline that a hypothetical classifier can achieve; this baseline is equal to proportionate size of the majority group. For example, if 88 of 100 documents are academic and 12 are news, the best strategy for a classifier is to guess *academic* every time; this hypothetical classifier will have an accuracy performance of 88%. This is what we call a *best guess* classifier.

The standard accuracy measure therefore under-rates the performance of a classifier that does better than the best guess and over-rates the performance of a classifier that does less well than best-guess. Suppose that one classifier identifies 88% of 100 documents correctly and another identifiers 90% of 100 correctly. There is only a 2% difference in performance by the two classifiers.

Accuracy gain (AG) achieves a more realistic measure of the performance of the classifiers by treating the baseline (or best-guess) performance as zero; the performance of a classifier that does better than best guess is represented with a positive number; the performance of a classifier that does less well than the baseline is represented as a negative number. Table 3 compares the measures of Accuracy and Accuracy Gain for a case where 88 documents belong to one class and 12 belong to another.

Classifier Performance (Baseline = 0.88)	
Accuracy (%)	Accuracy Gain (%)
86	-17
88	0
90	17
91	25
100	100

Table 3: Accuracy Computed as Accuracy Gain

With AG, a classifier that achieves only baseline Accuracy (88%) performs at 0%, i.e., no better or worse than it would have achieved with a best-guess strategy. But the classifier that achieves a two-point improvement (90%) over the baseline (88%) has a 17% Accuracy Gain. This more accurately reflects the performance of the classifier with the apparently small (2%) improvement.

Figure 1 shows how we calculate AG by re-scaling (or normalizing) the rigorous baseline to zero. The resulting fraction (AG) represents the improvement over the best guess procedure compared to the maximum possible improvement (100%).

AG=(Accuracy-Best Guess)/(100-Best Guess)
AG=(90-88)/(100-88)

Original Scale (Accuracy=90%)
Best Guess <------Accuracy---------------------->Perfect
88% 90% 100%

Re-Scale (Accuracy Gain=17%)
Best Guess <-----Accuracy Gain ---------------->Perfect
0% 17% 100%

Figure 1: Calculation of Accuracy Gain (AG)

Another advantage of the AG measure is that it allows consistent comparison of results for studies that use classes with different populations. All of the studies cited in Section 2 use the standard Accuracy measure; only half report the proportionate size of genre categories (Karlgren and Cutting [3], Kessler et al. [4], Ng et al. [7], Lee and Myaeng [8], Santini et al [10] and Santini [11], and zu Eissen and Stein [12]).

We therefore report our results using AG as the most rigorous performance measure and as a method that promotes comparability of future research.

5 RESULTS

Table 4 shows the impact of the different features (methods) for discriminating the three genres. As expected, the results for the one-against-one classification problems are better than one-against-many. Still, for both sets of problems, models

[3] See Rittman [1].

using adjective and adverb features outperform models containing nouns, verbs, or punctuation. For example, for academic vs. fiction, NN achieves an AG of 81.4%, higher than VB or Punc. But four features (SOA1; SOA2; SOA1+JJ3+RB2; All-JJ&RB) do better than these standard categories. This shows that adjectives and adverbs should be used as discriminating features for genre identification.

We also find that the highest AG for all six problems is achieved by the combination of many kinds of adjective and adverb features (All-JJ&RB); this row is highlighted in Table 4. For example, distinguishing academic writing from fiction, using this feature achieves an astonishing 98.8% of the possible gain in accuracy (AG). The same method applied to fiction vs. news and academic vs. news scores the second and third highest AG of 93.0% and 90.8%, respectively. The same is true for the harder problems: news vs. not-news (AG=13.5%), academic vs. not-academic (AG=10.6%) and an impressive AG of 52.5% for fiction vs. not-fiction.

Feature / Method	Problem					
	One-Against-One			One-Against-Many		
	Acad vs. Fict	Acad vs. News	Fict vs. News	Acad vs. Not-Acad	Fict vs. Not-Fict	News vs. Not-News
SOA1	91.6	75.6	69.8	-1.8	16.4	-1.5
SOA2	89.4	73.0	72.0	1.8	13.9	0.8
JJ	72.2	65.0	61.6	9.7	0.0	6.8
RB	77.6	41.8	88.4	0.0	-4.9	-0.8
SOA1+JJ3 +RB2	88.6	79.6	70.4	-2.7	14.8	2.3
All-JJ&RB	98.8	90.8	93.0	10.6	52.5	13.5
NN	81.4	68.6	83.0	-2.7	0.0	3.8
VB	72.2	69.0	76.0	-10.6	-8.2	0.0
Punc	76.8	20.4	83.4	0.0	28.7	-1.5

Table 4: Best Performing Models Using Adjectives and Adverbs Compared to Other Features

Table 4 also shows that for five of the six problems, the next best performance is achieved by vector models using the 30 SOAs (SOA1) (academic vs. fiction, AG=91.6%), the 110 adjectives and adverbs (SOA1+JJ3+RB2) (academic vs. news, AG=79.6%), or models using the simple count of all adjectives (JJ) (academic vs. not-academic, AG=9.7%; news vs. not-news, AG=6.8%) or all adverbs (RB) (fiction vs. news, AG=88.4%). For fiction vs. not-fiction, a model using the simple count of all punctuation achieves the second best result (AG=28.7%), compared to using all adjective and adverb features (All-JJ&RB) (AG=52.5%).

Another way to assess the performance of our methods is to see which choices of features produce an accuracy gain for all six classification problems. Table 4 shows that these methods include the count of all adjective and adverb features (All-JJ&RB), and the vector of the 30 SOAs (SOA2), although we see that AG for the hard problems of academic vs. not academic and news vs. not-news is only 1.8% and 0.8% respectively. Nevertheless, benchmark models using nouns, verbs, and punctuation do not achieve a positive AG for all six problems.

Furthermore, the model representing the 30 SOAs (SOA2) that yields a positive AG for all six problems contains only 11 to 19 unique words in the final discriminant model (Table 5). Fewer than 20 unique SOAs (SOA2) can do the work of thousands of words (All-JJ&RB).

Speaker-Oriented Adverbs (SOA2)		
Problem	**AG**	**Unique Words in Model ***
Acad vs. Fict	89.4	17
Acad vs. News	73.0	11
Fict vs. News	72.0	13
Acad vs. Not-Acad	1.8	13
Fict vs. Not-Fict	13.9	19
News vs. Not-News	0.8	16
* A unique word in a discriminant model can be represented as both a type and a token variable, or only as a type or a token		

Table 5: Number of Unique SOAs in Models Yielding Accuracy Gain for All Problems

We also assess classifier performance of vector models using combinations of the three sets of the 110 words (Appendix A, B, C). We find that models representing only the 30 SOAs are most effective for academic vs. fiction, fiction vs. news, and fiction vs. not-fiction (Table 6). When combined with trait adjectives and trait adverbs, performance improves slightly for academic vs. news and remains stable for fiction vs. not-fiction.

Feature / Method	Problem					
	One-Against-One			One-Against-Many		
	Acad vs. Fict	Acad vs. News	Fict vs. News	Acad vs. Not-Acad	Fict vs. Not-Fict	News vs. Not-News
SOA1	91.6	73.8	71.2	-3.5	14.8	-0.8
JJ3	68.4	48.6	46.0	-3.5	-5.7	1.5
RB2	47.8	21.6	38.2	-2.7	-1.6	-2.3
JJ3+RB2	71.8	51.8	48.8	-4.4	-0.8	2.3
SOA1+JJ3 +RB2	88.6	79.6	70.4	-2.7	14.8	2.3

Table 6: Contribution of Speaker-Oriented Adverbs, Trait Adjectives, and Trait Adverbs to Models

Classifier performance is slightly better than the best guess for news vs. not-news using trait adjectives and trait adverbs alone. However, performance is not effective for academic vs. not academic using any combination of the three sets of the

110 words. This suggests that, as a class, SOAs contribute more to vector models than do trait adjectives and trait adverbs, and none of the 110 words are effective for distinguishing academic from not-academic documents.

Finally, we assess the contribution of the 30 SOAs (Appendix C) as compared to the 44 trait adjectives (Appendix A) and the 36 trait adverbs (Appendix B) by ordering the relative contribution of these 110 words to our vector models. We rank the 110 words that we entered into the various models by assigning scores according to the contribution (weight) each word made to a model, and by giving credit for the number of models each word contributed to. This method evaluates the contribution of words to all possible models (though it does not show which words are best for discriminating between particular genres). We find that only 95 of the 110 words contributed to a vector model. These 95 words include the 30 SOAs, 40 of the 44 trait adjectives, and only 25 of the 36 trait adverbs (we indicate these words in Appendix A, B, and C). On average, SOAs made the greatest contribution to vector models, generally ranking higher than trait adjectives and trait adverbs. For example, half of the 30 SOAs (e.g., *maybe, generally, surely, necessarily, clearly, specifically, strangely, and seriously*) rank in the top 25% of most effective words, whereas only small numbers of the other classes occur above the same cutpoint (9 of 40 trait adjectives; e.g., *bad, moral, natural, characterless*, and *honest*), and only 3 of 25 trait adverbs: *fairly, badly*, and *naturally*). Clearly, SOAs contributed most significantly to our genre classification models.

This may be indicative of a relationship between narrator behavior (marked by the use of SOAs in text) and author intent (one of several distinguishing criteria of genre). It also shows that the use of a linguistically defined construct guides us directly to the essential feature of the statistical models. Indeed, a model representing only 30 SOAs (SOA1) is comparable to the best-performing model (All-JJ&RB) for academic vs. fiction (AG=91.6%) (Table 4). It is most difficult to distinguish the three genres from all other genres in the corpus as expected, although fiction vs. not-fiction is relatively distinct using this feature (SOA1) (AG=16.4%).

6 CONCLUSION AND FUTURE WORK

Motivated by research in psychology and linguistics, we demonstrate that using adjective and adverb features in discriminant models is generally superior to benchmark models containing nouns, verbs, or punctuation features. In particular, vector models representing only 110 words (SOAs, trait adjectives and trait adverbs) are comparable to models using the count of thousands of words. As a class, the 30 SOAs are generally more effective than the class of 44 trait adjectives and 36 trait adverbs for the classification experiments we conducted.

But in the long term, our specific results are less important than the evidence that our approach to systematically studying the contribution of adjectives and adverbs to genre identification provides useful clues about how expressions of affect can be recognized by computer systems and how this information can be used for any application that depends on accurate identification of characteristics of affective language. Accuracy Gain rigorously measures the contribution of features to classification problems.

We recommend that our principles and methods be applied to (a) solving problems in other applications, (b) using other corpora, and (c) finding other features. Other applications include author identification, detection of subjectivity versus objectivity, classification of texts for question-answering, natural language generation, detection of customer review opinions in business environments, detection of personality traits in text, and detection of people who might be susceptible to certain beliefs. Other corpora include weblogs, email logs, and chat room logs. Possible sources of features include the domains of stylistics, communication, journalism, content analysis, and political discourse.

REFERENCES

[1] R. Rittman. *Automatic Discrimination of Genres: The Role of Adjectives and Adverbs as Suggested by Linguistics and Psychology*. Ph.D. Dissertation, Rutgers, The State University of New Jersey, New Brunswick, NJ. (2007). https://www.scils.rutgers.edu/~rritt/Rittman_dissertation_20070427.pdf

[2] D. Lee. Genres, Registers, Text Types, Domains and Styles: Clarifying the Concepts and Navigating a Path Through the BNC Jungle. *Language, Learning, and Technology*, 5(3):37-72 (2001).

[3] J. Karlgren and D. Cutting. Recognizing Text Genres with Simple Metrics Using Discriminant Analysis. *Proceedings of the 15th. International Conference on Computational Linguistics*, Coling 94, Kyoto. 1071-1075. (1994).

[4] B. Kessler, G. Nunberg, and H. Schutze. Automatic Detection of Text Genre. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 32 - 38. (1997).

[5] M. Wolters and M. Kirsten. Exploring the Use of Linguistic Features in Domain and Genre Classification. *9th Conference of the European Chapter of the Association for Computational Linguistics*, Bergen, Norway, 142-149. (1999).

[6] E. Stamatatos, N. Fakotakis, and G. Kokkinakis. Text Genre Detection Using Common Word Frequencies. *18th International Conference on Computational Linguistics* (COLING 2000), Luxembourg, Vol. 2, 808-814. (2000).

[7] K.B. Ng, S. Rieh, and P. Kantor. Signal Detection Methods and Discriminant Analysis Applied to Categorization of Newspaper and Government Documents: A Preliminary Study. In *Proceedings of Annual Conference of American Society for Information Science*, pp. 227-236. (2000).

[8] Y. Lee and S. Myaeng. Text Genre Classification with Genre-Revealing and Subject-Revealing Features. *Proceedings of the 25th Annual International ACM SIGIR*, 145-150. (2002).

[9] C. Lim, K. Lee and G. Kim. Multiple Sets of Features for Automatic Genre Classification of Web Documents. *Information Processing and Management*, 41, 1263–1276. (2005).

[10] M. Santini, R. Power and R. Evans. Implementing a Characterization of Genre for Automatic Identification of Web Pages. *Proceedings of 21st International Conference on Computational Linguistics* and *44th Annual Meeting of the Association for Computational Linguistics*, 699-706. (2006).

[11] M. Santini. Some Issues in Automatic Genre Classification of Web Pages. *Proceedings of JADT 2006: 8es Journées Internationales d'Analyse statistique des Données Textuelles*. (2006).

[12] S. M. zu Eissen and B. Stein. Genre Classification of Web Pages: User Study and Feasibility Analysis. *Proceedings of the 27th Annual German Conference on Artificial Intelligence (KI 04)*, 256-269. (2004).

[13] A. Finn and N. Kushmerick. Learning to Classify Documents According to Genre. *Journal of the American Society for Information Science and Technology*, 57 (11), 1506-1518. (2006).
[14] R. Bruce and J. Wiebe. Recognizing Subjectivity: A Case Study in Manual Tagging. *Natural Language Engineering* 5(2):187-205. (1999).
[15] J. Wiebe. Learning Subjective Adjectives from Corpora. *Proceedings of the 17th National Conference on Artificial Intelligence* (AAAI-2000), Austin. (2000).
[16] J. Wiebe, R. Bruce, and T. O'Hara. Development and Use of a Gold Standard Data Set for Subjectivity Classifications. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, College Park: University of Maryland, 246-253. (1999).
[17] R. Rittman, N. Wacholder, P. Kantor, K.B. Ng, T. Strzalkowski, B. Bai, et al. Adjectives as Indicators of Subjectivity in Documents. *Proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology*, 349-359. (2004).
[18] L. Goldberg. The Development of Markers for the Big-Five Factor Structure. *Psychological Assessment*, 4:26-42. (1992).
[19] S. Nowson, J. Oberlander, and A. Gill. Weblogs, Genres and Individual Differences. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, Stresa, Italy. (2005).
[20] S. Argamon, S. Dhawle, M. Koppel, and J.W. Pennebaker. Lexical Predictors of Personality Type. *Proceedings of the Classification Society of North America*, St. Louis. (2005).
[21] F. Mairesse, M.A. Walker, M.R. Mehl, R.K. Moore. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research*, 30, 457-500. (2007).
[22] D. Peabody and B. De Raad. The Substantive Nature of Psycholexical Personality Factors: A Comparison Across Languages. *Journal of Personality and Social Psychology*, 83(4):983-997. (2002).
[23] R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik. *A Comprehensive Grammar of the English Language*. London: Longman. (1985).
[24] A. Mittwoch. How to Refer to One's Own Words: Speech Act Modifying Adverbials and the Performative Analysis. *Journal of Linguistics*, 13:177-189. (1977).
[25] R. Jackendoff. Adverbs. In R. Jackendoff. *Semantic Interpretation in Generative Grammar* (pp. 47-107). Cambridge, MA: MIT Press. (1972).
[26] T. Ernst. The Semantics of Predicational Adverbs, and The Scopal Basis of Adverb Licensing. In T. Ernst. *The Syntax of Adjuncts* (pp. 41-91, 92-148). New York: Cambridge University Press. (2002).
[27] BNC2. British National Corpus, World Edition. http://www.natcorp.ox.ac.uk/ (2001).
[28] D. Lee. BNC World Index. http://personal.cityu.edu.hk/~davidlee/devotedtocorpora/home/BNC_WORLD_INDEX.ZIP (2003).
[29] D. Lee. David Lee's Genre Classification Scheme. http://homepage.mac.com/bncweb/manual/genres.html (n.d.).
[30] M. Davies. VIEW: Variation in English Words and Phrases. http://view.byu.edu/ (n.d.).
[31] J. Wiebe. [Learning Subjective Adjectives from Corpora]. Unpublished list of subjective adjectives. http://www.cs.pitt.edu/~wiebe/pubs/aaai00/adjsMPQA (2000).
[32] I. Witten and E. Frank. *Data Mining*. San Diego: Academic Press. (2000).

APPENDIX A: 44 Trait Adjectives (JJ3)

avaricious *	bad *	biased *	calculating *	characterless *
decent *	deceptive *	dishonest *	disloyal *	ethical *
fair *	faithful *	frank *	honest *	hypocritical *
insincere *	intriguing *	just *	loyal *	lustful *
lying *	malicious *	materialistic *	mercenary *	moral *
natural *	noble *	perfidious	pharisaical	principled *
rapacious *	righteous *	sincere *	trustworthy *	truthful *
underhanded	unfaithful *	unreliable *	unscrupulous *	untruthful *
upright *	venal	virtuous *	vulgar *	
* Indicates 40 words that contributed to a vector model				

APPENDIX B: 36 Trait Adverbs (RB2)

avariciously	badly *	calculatingly *	decently *	deceptively *
dishonestly *	disloyally *	ethically *	fairly *	faithfully
hypocritically *	insincerely *	intriguingly *	justly *	loyally *
lustfully *	maliciously *	materialistically *	morally *	naturally *
nobly *	perfidiously *	pharisaically	rapaciously	righteously *
sincerely *	truthfully *	underhandedly	unfaithfully	unreliably *
unscrupulously	untruthfully	uprightly	valiantly	virtuously *
vulgarly				
* Indicates 25 words that contributed to a vector model				

APPENDIX C: 30Speaker-Oriented Adverbs (SOAs)

amazingly	briefly	candidly	certainly	clearly
confidently	curiously	definitely	frankly	generally
honestly	ideally	luckily	maybe	necessarily
normally	obviously	oddly	possibly	predictably
preferably	probably	roughly	seriously	simply
specifically	strangely	surely	surprisingly	unfortunately
Note: All SOAs contributed to a vector model				