Stylistic Text Classification Using Functional Lexical Features

Shlomo Argamon[1] Casey Whitelaw[2*] Paul Chase[1] Sushant Dhawle[1]
Sobhan Raj Hota[1] Navendu Garg[1] Shlomo Levitan[1]

[1]Linguistic Cognition Lab, Department of Computer Science
Illinois Institute of Technology
10 W. 31st Street, Chicago, IL 60616, USA

[2]School of Information Technologies, University of Sydney
Sydney, NSW 00000, Australia

July 20, 2005

Abstract

Most text analysis and retrieval work to date has focused on determining the topic of a text, what it is about. However, a text also contains much useful information in its *style*, or how it is written. This includes information about its author, its purpose, feelings it is meant to evoke, and more. This paper addresses the problem of classifying texts by style (along several different dimensions), developing a new type of lexical feature based on taxonomies of various semantic *functions* of different lexical items (words or phrases). We show the usefulness of such features for text classification by author, author personality, gender of literary characters, sentiment (positive/negative feeling), and scientific rhetorical styles. We further show how the use of such functional features aids in gaining insight about stylistic differences between texts.

[*]Casey Whitelaw was a visiting scholar at the IIT Linguistic Cognition Laboratory during November 2004.

1 Introduction

A common goal in automated text analysis is to gain an understanding or summary of the topic, or topics, covered in the text. This may involve information extraction into frame-based semantic representations (Hammer, Garcia-Molina, Cho, Crespo, & Aranha, 1997), text clustering and categorization (Sebastiani, 2002; Kehagias, Petridis, Kaburlasos, & Fragkou, 2003), or simply retrieval of topic-relevant documents by one of a variety of heuristics (Salton & McGill, 1983; Ponte & Croft, 1998; Grossman & Frieder, 1998). However, it is now recognized that much more than such 'objective' representations of the information in a text are needed to properly support users in interactive retrieval tasks (Belkin, 1993; Chen, Magoulas, & Dimakopoulos, 2005). Dealing with issues of 'information quality' (Tang, Ng, Strzalkowski, & Kantor, 2003) and 'authority' (Fritch & Cromwell, 2001) have been identified is important for more effective user support. In this paper, we examine another dimension: how to extract useful *stylistic* information from texts.

We view the full meaning of a text is much more than just the topic it describes or represents. Textual meaning, more broadly construed, includes also aspects such as: *affect* (what feeling is conveyed by the text?), *genre* (in what community of discourse does the text function?), *register* (what is the function of the text as a whole?), and *personality* (what sort of person, or who specifically, wrote the text?). These aspects of meaning are captured by the text's *style* of writing, which may be roughly defined as how the author chose to express her topic, from among a very large space of possible ways of doing so. We contrast, therefore, the **how** of a text (style) from the **what** (topic).

Immediate applications of stylistic text analysis include authorship attribution and profiling (Mosteller & Wallace, 1964; Burrows, 1987; Kjell, Woods, & Frieder, 1994; Baayen, Halteren, & Tweedie, 1996; de Vel, Corney, Anderson, & G.Mohay, 2002; Argamon, Koppel, Fine, & Shimony, 2003; Stamatatos, Fakotakis, & Kokkinakis, 2000; McEnery & Oakes, 2000), genre-based text classification and retrieval (Karlgren & Cutting, 1994; Kessler, Nunberg, & Schütze, 1997; Finn, Kushmerick, & Smyth, 2002), sentiment analysis (P. Turney & Littman, 2002; Pang, Lee, & Vaithyanathan, 2002), and spam and scam filtering (Androutsopoulos, Koutsias, Chandrinos, Paliouras, & Spyropoulos, 2000; Kushmerick, 1999; Patrick, 2004). Other applications include criminal and national security forensics (Chaski, 1999; McMenamin, 2002), mining of customer

feedback (Berry & Linoff, 1997; McKinney, Yoon, & Zahedi, 2002), and aiding humanities scholarship (Matthews & Merriam, 1997; Holmes, 1998; Hoover, 2002). As the sheer quantity of texts available on every given topic grows exponentially, the need for automated extraction of more dimensions of meaning is becoming acute.

This paper addresses a key problem for stylistic text categorization methods, to wit, what features of a text to use for modeling style. While topic-based text categorization can get quite far by using models based on "bags of content words", style is somewhat more elusive. We start from the intuitive notion that style is indicated by features that indicate the author's choice of one mode of expression from among a set of equivalent modes for a given content. At the surface level, this may be expressed by a wide variety of possible features of a text: choice of particular words, syntactic structures, discourse strategy, or all of the above and more. The underlying causes of such surface variation are similarly heterogeneous, including the genre, register, or purpose of the text, as well as the educational background, social status, and personality of the author and audience. What all these dimensions of variation have in common, though, is an independence from the 'topic' or 'content' of the text, which may be considered to be those objects and events that it refers to (as well as their properties and relations as described in the text). We may thus provisionally define the *stylistic meaning* of a text to be those aspects of its meaning that are *non-denotational*, i.e., independent of the objects and events to which the text refers.

Most computational stylistics work to date has been based on hand-selected sets of content-independent features such as function words (Mosteller & Wallace, 1964; Matthews & Merriam, 1997; Tweedie, Singh, & Holmes, 1996), parts-of-speech and syntactic structures (Stamatatos et al., 2000), or clause/sentence complexity measures (Yule, 1944; Vel, 2000) (also see the survey in (Karlgren, 2000)). New developments in machine learning and computational linguistics have enabled larger numbers of features to be generated for stylistic analysis, in almost no case is there strong theoretical motivatation behind input feature sets, where features would be devised to be linguistically related to style issues. Rather the general methodology that has developed is to find as large a set of topic-independent textual features as possible and use them as input to a generic learning algorithm (preferably one resistant to overfitting, and possibly including feature selection). Indeed, some interesting and effective feature sets have been found in this way (such as (Karlgren, 2000; Koppel, Akiva, & Dagan, 2003); function words have also proven to

be surprisingly effective on their own (McEnery & Oakes, 2000; Argamon et al., 2003; Argamon & Levitan, 2005)). Nevertheless, we contend that without a firm basis in a linguistic theory of meaning, we are unlikely to gain any true insight into the nature of the stylistic dimension(s) under study. Proper choice of features should also boost classification accuracy.

Our goal, therefore, is to find a computationally tractable formulation of linguistically well-motivated features which permit text classification based on variation in non-denotational meanings. We describe here a framework and methodology for constructing a lexicon using attribute-value taxonomies based on principles of Systemic Functional Grammar (SFG) (Halliday, 1994), which we find to be useful for this purpose. In particular, SFG explicitly recognizes and represents *non-denotational* meaning as part of the general grammar, which makes it particularly applicable to stylistical problems.

Our current system does no complex parsing, relying instead on mainly lexical features of the text. While more sophisticated processing will likely lead to improved results in the future, we believe that simpler methods should be fully explored first. The text analysis methodology described in this paper comprises four steps:

1. Tokenize texts and assign part-of-speech tags;
2. Extract instances of lexical units from each text, as specified in the lexicon (with some, but not complete, disambiguation);
3. Compute relative frequencies of semantic attribute values for each text, giving an overall "feature vector" describing the text;
4. Use machine learning to construct discrimination models for stylistic text classification tasks.

Using this methodology, we show below (Sec. 6) how functional lexical features can improve stylistic classification results for a variety of stylistic tasks:

Authorship attribution: Determining who (out of a given list of candidates) wrote a given chapter of a literary work;

Gender attribution: Determining whether a speech in one of Shakespeare's plays is said by a male or a female character;

Personality typing: Determining if the author of a short self-reflective text is neurotic or not;

Sentiment analysis: Determining if a movie review is positive or negative, based only on the text; and

Scientific rhetoric: Determining if two similar scientific fields (Geology and Paleontology) differ in their reasoning and argumentation structure, by analyzing peer-reviewed journal articles.

Paper outline

The remainder of the paper describes the overall design of the system and lexicon (Sec. 2), the semantic attributes currently used in the lexicon (Sec. 3), how to construct such a lexicon (Sec. 4), our experimental methodology (Sec. 5), and experimental results for several stylistic text classification tasks, demonstrating the usefulness of our functional lexical features (Sec. 6).

2 Overall Design

2.1 ATMan

The core of our style analysis system is ATMan[1], in which input texts are linguistically processed, creating 'annotated texts' (*atexts*), stored in a relational database. The system is written in Java, for portability.

An atext comprises several tables in the database, as follows. The main component of an atext is a sequence of *tokens*, each corresponding to a word, number, or punctuation mark, and each labeled with a set of *attributes* such as the token's position (in the text, in its sentence), part-of-speech (singular noun, past-tense verb, etc.), capitalization (lowercase, capitalized, all-uppercase, etc.), length in characters, and so forth. Raw text is converted into an atext in the database by an *import method* which uses a library of tokenizers and token analysis methods. Different import methods are defined for different input file formats (Usenet articles, SGML-tagged corpus files, etc.).

An atext also has a set of linguistic *units* that have been extracted based on a semantic lexicon of words and phrases, each corresponding to a short sequence of tokens in the text and described

[1] **A**nnotated **T**ext **Man**ager; beta-release anticipated in late 2005.

by a set of semantic attribute-value pairs (as described further below). Metadata about the text (its title, source, author, etc.) are also stored in the atext.

A separate data element in ATMan is the lexicon itself, used for semantic unit extraction, described in detail in Section 2.2 below.

For use in machine learning, ATMan outputs ARFF files (in Weka (Witten & Frank, 2000) format) comprising a list of labeled numeric vectors, each corresponding to an atext. Each vector element represents the relative frequency of some feature-value in the atext, conditional on some other possible feature-value (see Section 3), with labels indicating possible classes for text classification learning.

2.2 Lexicon design

The lexicon comprises a set of *lexical items* (words or phrases), each described by one or more *lexical entries*. Multiplicity of entries for a lexical item corresponds to ambiguity of meaning. Each lexical entry consists of both *constraints* on when the entry is applicable to the lexical item, and a set of *attribute values* describing syntactic and semantic properties. Currently, the only type of constraint is expressed as a set of allowed part-of-speech sequences, such that the entry is only instantiated for the lexical item if it is tagged with an allowed part-of-speech sequence (recall that lexical items may be multiword phrases). Each attribute is the name of a system network (Sec. 3 below), and is assigned a set of values, each corresponding to an option in that system network. All of the values are assigned conjunctively to that attribute in that entry.

This general structure is unexceptional; it is from the choice of attributes and the semantic organization of their possible values that gives the approach any power. We note again that our goal is *not* syntactic parsing, and hence we rely as much as possible on locally-computable properties of the text, specifically part-of-speech tags, to constrain interpretation. Furthermore, as the final goal of this processing is to compute an overall statistical representation of the document (as a vector of numeric feature frequencies), some local ambiguity in interpretation can be tolerated.

3 Functional Lexical Attributes

This section gives an overview of the attributes we have found useful to date for stylistic text classification; more detail may be found in the Appendix. Our work is based on the theory of Systemic Functional Grammar (SFG), a functional approach to linguistic analysis (Halliday, 1994). SFG models languages as a system of choices of meanings to represent in language (Matthiessen, 1995), and so all lexical and structural choices are represented in terms of their semantic functions. The theory has been applied to natural language processing in several contexts since the 1960s, but has been used most widely in text generation (Matthiessen & Bateman, 1991; Teich, 1995), due to the difficulty of full parsing in the theory.

SFG construes language as a set of interlocking choices for expressing meanings, with more general choices constraining the possible specific choices. A simple example in English:

> If a pronoun is to be used, it may refer either to one of the discourse participants, or to a third party;
>
> - If to one of the participants, it may refer to the speaker (*I, me*), the speaker-plus-others (*we, us*), or the hearer (*you*);
> - If to a third party, it may refer either to one individual or to many (*they, them*);
>   - If to a single individual, it may refer to a conscious individual or to a non-conscious individual (*it*);
>   - If to a single conscious individual, it may refer to a male (*he, him*) or to a female (*she, her*);
>
> and so forth...

Note that a choice at one level may open up further choices at other levels, choices that are not open otherwise; e.g., English does not allow a pronoun to distinguish between pluralities of conscious or non-conscious individuals. Furthermore, any specific choice of lexical item or syntactic structure is determined by choices from multiple systems at once, as the choice between "I" and "me" is determined by the independent choice governing the pronoun's syntactic role as either a subject or an object.

Thus a *system* defines a set of *options* for meanings to be expressed. Each (non-root) system has an *entry condition*, a propositional formula of options from other systems, denoting when that system is possible. Each option gives constraints (lexical, morphological, or syntactic) on utterances that express the option. Options (or logical combinations thereof) may serve as entry conditions for more specific systems. While some systems, as in the example above, are *disjunctive* such that exactly one of their options must be chosen, others are *conjunctive* in that all of their options must be chosen—this enables combinatorial possibilities. For example, modal verbs (such as 'may', 'might', or 'must') choose options from multiple systems, including "Modality Type" (likelihood, frequency, obligation, etc.) and "Modality Value" (median, high, low).

In our current work, each lexical entry in the lexicon is assigned a value for each of a set of *semantic lexical attributes* from the options in associated *system networks*. Each such network has a unique root, and we allow entry conditions to be only single options or conjunctions of options[2]. More formally, each system network in this conception is a directed acyclic AND/OR graph, whose nodes are systems and whose directed arcs are options. An option O_1 is a *child* of option O_2 if O_1's destination node is O_2's source node; descendants and ancestors in the graph are defined in the straightforward manner. If option O_1 is chosen and it leads into a disjunctive node, then exactly one of its children must also be chosen; if it leads into a conjunctive node, then all of its children must also be chosen. Note that if an option is chosen, all of its ancestors are also chosen.

As noted above, each lexical entry is a frame comprising a set of attribute values, where each attribute is the name of a system network, and each value is an option (or set of noncontradictory options) in the system network. Documents are represented by numeric feature vectors, where each feature is the relative frequency of some option O_1 with respect to some other option O_2. Given an atext d, define $N_d(O_1)$ to be the number of units in d with value O_1, similarly $N_d(O_1, O_2)$ to be the number with both O_1 and O_2. Then the *relative frequency of* O_1 *with respect to* O_2 is defined as

$$RF_d(O_1|O_2) = \frac{N_d(O_1, O_2)}{N_d(O_2)}$$

For example, the frequency of sibling options relative to their shared parent allows direct comparison of how different texts prefer to express the parent via its different options. Alternatively, the

[2]See (Matthiessen, 1995) for a discussion of the full SFG grammar representation (allowing disjunction in entry conditions) which we simplify for computational ease.

frequency of options relative to a system network root enables a more global comparison of what types of meanings (with a given system) are expressed in a document. Other kinds of relative frequency features can be useful as well, as discussed below.

The remainder of this section fleshes out the main system networks which we use here for computational analysis of textual style. They can be divided into three categories, denoting the general 'stylistic goals' that these textual features relate to: *Cohesion*, referring to how a text is constructed to 'hang together', *Assessment*, meaning how a text construes propositions as statements of belief, obligation, or necessity, contextualizing them in the larger discourse, and *Appraisal*, or how the text adjudges the quality of various referred-to objects or events. Note that the system networks we use are the result of decades of research on textual analysis within the SFG community, and are not *ad hoc* inventions for our particular purposes.

3.1 Cohesion

Cohesion refers to linguistic resources that enable language to connect to its larger context, both textual and extratextual (Halliday & Hasan, 1976). Such resources include a wide variety of referential modalities (pronominal reference, deictic expressions, ellipsis, and more), as well as lexical repetition and variation, and different ways of linking clauses together. How an author uses these various cohesive resources is an indication of how the author organizes concepts and relates them to each other. Within cohesion, our current computational work considers just types of conjunctions, for feasibility in automated extraction. Coreference resolution, for example, is a very difficult unsolved problem.

Words and phrases that conjoin clauses (such as 'and', 'while', and 'in other words') are organized in SFG in the CONJUNCTION system network. Types of CONJUNCTION serve to link a clause with its textual context, by denoting how the given clause expands on some aspect of its preceding context (Matthiessen, 1995, p. 519–528). The three top-level options of CONJUNCTION are Elaboration, Extension, and Enhancement:

- Elaboration: Deepening the content in its context by exemplification or refocusing.
- Extension: Adding new related information, perhaps contrasting with the current information.

- Enhancement: Qualifying the context by circumstance or logical connection.

A more detailed description of the CONJUNCTION taxonomy is given in the Appendix.

3.2 Assessment

Generally speaking, *assessment* may be defined as "contextual qualification of the epistemic or rhetorical status of events or propositions represented in a text". Examples include assessment of the likelihood of a proposition, the typicality of an event, the desirability of some fact, or its scope of validity. Two important systems in SFG that address assessment are MODALITY, enabling expression of typicality and necessity of some fact or event, and COMMENT, enabling assessment of the writer's attitude or stance towards to an assertion in the text.

The system of MODALITY enables writers to qualify events or entities in the text according to their likelihood, typicality, or necessity. Syntactically, MODALITY may be realized in a text through a modal verb (e.g., 'can', 'might', 'should', 'must'), an adverbial adjunct (e.g., 'probably', 'preferably'), or use of a projective clause (e.g., "I think that...", "It is necessary that..."). Each expression of MODALITY has a value for each of four attributes (see the discussion in the appendix for more detail):

- Type: What kind of modality is being expressed?
  - Modalization: How 'typical' is it? (*probably*, *seldom*)
  - Modulation: How 'necessary' is it? (*ought to*, *allowable*)
- Value: What degree of the relevant modality scale is being averred?
  - Median: The 'normal' amount. (*likely*, *usually*)
  - Outer: An extreme (either high or low) amount. (*maybe*, *always*)
- Orientation: Relation of the modality expressed to the speaker/writer.
  - Objective: Modality expressed irrespective of the speaker/writer. (*maybe*, *always*)
  - Subjective: Modality expressed relative to the speaker/writer. (*We think...*, *I require...*)
- Manifestation: How is the modal assessment related to the event being assessed?
  - Implicit: Modality realized 'in-line' by an adjunct or modal auxiliary. (*preferably...*, *maybe..*)

– Explicit: Modality realized by a projective verb, with the nested clause being assessed. (*It is preferable...*, *It is possible..*)

The system of COMMENT provides a resource for the writer to 'comment' on the status of a message with respect to textual and interactive context in a discourse. Comments are usually realized as adjuncts in a clause and may appear initially, medially, or finally. We use the eight categories of COMMENT listed by Matthiessen (1995): *Admissive*, message is an admission (e.g., '*we concur...*'), *Assertive*, emphasis of reliability (e.g., '*Certainly...*'), *Desiderative*, desirability of the content (e.g., '*Unfortunately...*') , *Evaluative*, judgment of the actors involved (e.g., '*Sensibly...*'), *Predictive*, coherence with predictions (e.g., '*As expected...*'), *Presumptive*, dependence on other assumptions (e.g., '*I suppose...*'), *Tentative*, assessing the message as tentative (e.g., '*Tentatively...*') , and *Validative*, assessing scope of validity (e.g., '*In...*eral...").

3.3 Appraisal

Finally, *appraisal* denotes how language is used to adopt or express an attitude of some kind towards some target (Martin & White, 2005). For example, in "I found the movie quite monotonous", the speaker adopts a negative *Attitude* ("monotonous") towards "the movie" (the *appraised object*). Note that attitudes come in different types; for example, 'monotonous' describes an inherent quality of the appraised object, while 'loathed' would describe an emotional reaction of the writer. The overall type and orientation of appraisal expressed in the text about an object gives a picture of how the writer wishes the reader to view it (modulo sarcasm, of course). To date, we have developed a lexicon for appraisal adjectives as well as relevant modifiers (such as 'very' or 'sort of'). The two main attributes of appraisal, as used in this work, are Attitude, giving the kind of appraisal being expressed, and Orientation, giving whether the appraisal is *positive* (good, beautiful, nice) or *negative* (bad, ugly, evil). (There are also other attributes of appraisal, as discussed in the appendix.) The three main types of Attitude are: *affect*, relating to the speaker/writers emotional state (e.g., 'happy', 'sad'), *appreciation*, expressing evaluation of supposed intrinsic qualities of an object (e.g., 'tall', 'complex'), and *judgment*, expressing social evaluation (e.g., 'brave', 'cowardly'). More detail on the appraisal taxonomy as used in this work is given in the appendix.

4 Constructing the lexicon

Lexicons in each system network described above were constructed using a semi-automated technique to find relevant terms and assign them appropriate attribute values. In each case, we started with *seed terms* taken from example words and phrases given for various combinations of system options in standard SFG references: Halliday's introduction to SFG (Halliday, 1994), Matthiessen's grammar of modern English (Matthiessen, 1995), and Martin and White's appraisal theory (Martin & White, 2005). Candidate expansions for each seed term were generated from multiple resources—WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) and from two online thesauri (`http://m-w.com` and `http://thesaurus.com`). In WordNet, the members of each synset (set of synonyms) were taken as the related set; similarly, synonym and related word lists were taken from each thesaurus. Candidates were accepted only with the same part of speech as a seed term.

A list was generated, for each main category, of all such candidate terms, and they were then ranked by frequency of occurrence in the candidate list (total number of seed term/resource pairs generating that candidate). This provided a coarse ranking of relevance, enabling more efficient manual filtering. Uncommon words, unrelated words, or words arising from an incorrect sense of a seed term will tend to be ranked lower in the candidate list than those related to more of the seed terms and are present in more of the resources. As well as increasing coverage, using multiple thesauri allows for more confidence votes and in practice increases the utility of the ranking.

Each ranked list was manually inspected to produce the final set of terms used. In practice, terms with low confidence were automatically discarded, reducing the amount of manual work required.

5 Methodology

5.1 Feature sets

We empirically evaluated the use of functional lexical features for stylistic classification by applying them as well as standard function words. The following section presents results for a variety of stylistic classification tasks, using the following methodology (applied to a different corpus in each

case). All documents in each corpus were processed into numeric feature vectors using various combinations of the following feature sets (refer to Section 3):

FW: Features are the relative frequencies of a set FW of 675 function words, with each such feature (for a given word $w \in FW$) defined as:

$$\frac{\text{count}(w)}{\sum_{w' \in FW} \text{count}(w')}$$

Con: Each feature is the relative frequency (RF_d) of a node in the Conjunction system (Fig. 8) with respect to its parent.

Mod: This feature set consists of the union of two related feature sets:

- For each node in each Modality system (Type, Value, Orientation, and Manifestation; cf. Fig. 9), the relative frequency (RF_d) of the node with respect to its parent;
- For each pair of nodes in different Modality systems (e.g., Type and Value), the relative frequency (RF_d) of terms labelled by both nodes with respect to the conjunction of their parents.

Com: This set consists of the relative frequency (RF_d) of each node in the Comment system (Sec. A.3) with respect to its parent.

Att: This feature set comprises, for each node in the Attitude system, the relative frequency (RF_d) of the node with respect to its parent;

App: This feature set comprises Att, as well as, for each node n in Attitude, both $RF_d(\text{Positive}|n)$ and $RF_d(\text{Negative}|n)$.

Combinations of these feature sets (amounting to concatenating the relevant feature vectors) were also considered (termed, e.g., Con+Mod, denoting the union of Con and Mod); as well, in some cases (detailed below) per-token frequencies were used and compared to the relative frequency features described here.

5.2 Machine learning

In each experiment Weka's (Witten & Frank, 2000) implementation of the SMO learning algorithm (Platt, 1998) with a linear kernel was used for learning classification models; for the multiclass problems, a simple one-vs-all strategy was used to generalize the binary SMO learner for multiple output classes. Except where otherwise noted, 10-fold cross-validation was used throughout to estimate out-of-training classification accuracy.

5.3 Feature analysis

In many cases, as we shall see, examining the most important features for stylistic classification can give useful insights. The classification importance of each feature is taken to be represented by the magnitude of its weight in the linear model constructed by SMO. To make explicit the relationship that the functional features indicating each of two document classes give us, we take the top features indicating each class and find all *oppositions* they give, where an opposition is a pair of relative frequencies features, one of which indicates one class and the other indicates the other class, where the features' conditioning events are identical and their conditioned events are sibling nodes in some systemic taxonomy. For example, if CONJUNCTION/Extension (i.e., RF_d(Extension|CONJUNCTION)) is indicative of class A and CONJUNCTION/Enhancement of class B, we would have the opposition:

Condition	**Class A**	**Class B**
CONJUNCTION	Extension	Enhancement

A more complex example is where class A is indicated by high values of

$$RF_d(\text{Median}|\text{VALUE,MODALITY TYPE/Modalization})$$

and class B by high values of

$$RF_d(\text{Low}|\text{VALUE,MODALITY TYPE/Modalization})$$

In this case, the conditioning event is the conjunction of two nodes, one of which is the shared parent of the conditioned events. This gives the opposition:

Condition	Class A	Class B
MODALITY TYPE/Modalization:VALUE	Median	Low

In this case, when a text in Class A expresses Modalization (typicality of an event or proposition), it prefers to express Median (i.e., non-extreme) values, whereas in similar situations, Class B prefers to express Low values. This may indicate that texts in Class A tend to be more cautious, not expressing even unexceptional statements as absolute fact (saying "he likely went home" rather than "he went home"), while texts in Class B might only explicitly express Modalization when it is particularly low (saying "he went home" in the last case, but "she might have wanted him to stay", if the conclusion is uncertain). Interpretation will depend, of course, on the particular types of texts under consideration.

The oppositions given by such analysis give direct information about linguistic differences between two document classes, in that the two classes have differing preferences about how to express the conditioning event. In the first example above, Class A prefers to conjoin items by Expansion, indicating a higher density of more-or-less independent information units, whereas Class B prefers conjoining items by Enhancements, indicating a more closely focused structure dealing with a smaller number of independent information units.

6 Experimental Results

To validate the methodology of using functional lexical features for stylistic classification, we ran a experiments on a number of different stylistic classification tasks, showing that (a) functional lexical features can aid classification, and (b) in many cases analyzing indicative features can give insight into underlying phenomena.

6.1 Authorship identification

Authorship attribution, the problem of determining who wrote an anonymous text, is perhaps the most classic stylistic text classification task. Ever since the influential work of Mosteller and Wallace (1964) on the authorship of the Federalist Papers function words, those with grammatical function (such as 'the', 'and', 'for'), have proven remarkably resilient for this task, even though many other potentially useful features have been suggested. The intuition behind the utility of

Table 1: The authorship attribution corpus, comprising the chapters in a set of 20 nineteenth-century novels.

Author	Book	# Chapters	Avg. Words
Cather	My Antonia	45	1826
	Song of the Lark	60	2581
	The Professor's House	28	2172
Conrad	Lord Jim	45	2913
	The Nigger of the Narcissus	5	10592
Hardy	Jude the Obscure	53	2765
	The Mayor of Casterbridge	45	2615
	Tess of the d'Urbervilles	58	2605
James	The Europeans	12	5003
	The Ambassadors	36	4584
Kipling	The Jungle Book	13	3980
	Kim	15	7167
Lewis	Babbit	34	3693
	Main Street	34	4994
	Our Mr. Wrenn	19	4126
London	The Call of The Wild	7	4589
	The Sea Wolf	39	2739
	White Fang	25	2917
Wells	The Invisible Man	28	1756
	The War Of The Worlds	27	2241

function words for stylistic attribution is as follows. Due to their high frequency in the language and highly grammaticalized roles, function words are very unlikely to be subject to conscious control by the author. At the same time, the frequencies of different function words vary greatly across different authors and genres of text - hence the expectation that modeling the interdependence of different function word frequencies with style will result in effective attribution. However, the highly reductionistic nature of such features seems unsatisfying, as they rarely give good insight into underlying stylistic issues. We suggest here that some of the systemic functional features developed in this work may both aid in accurate authorship attribution as well as give some insight.

6.1.1 The corpus

The corpus for this evaluation (see Table 1) was constructed from twenty nineteenth century novels by eight different authors (those used in Hoover's (2002) recent authorship study). Each novel was divided into individual chapters, each of which was considered as a separate example for learning and classification.

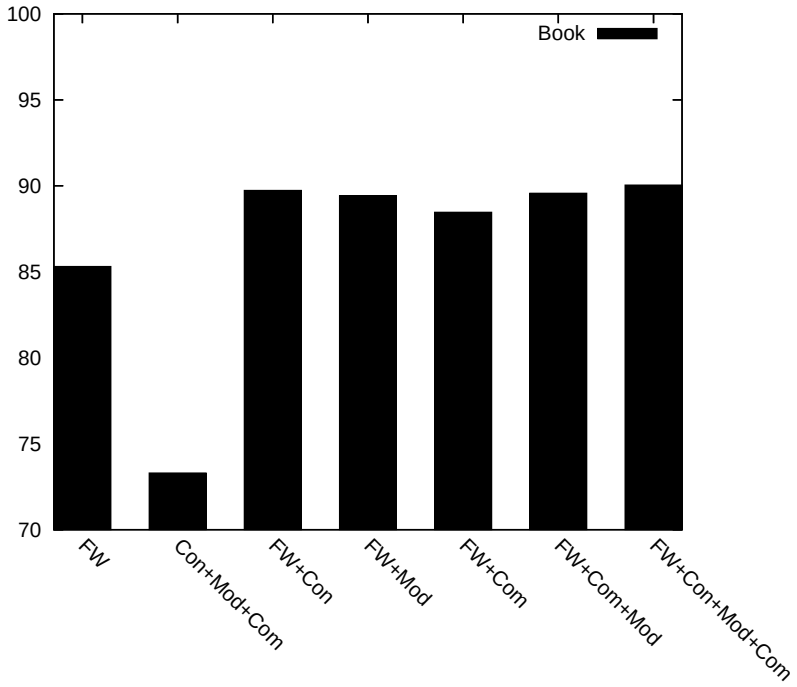

Figure 1: 10-fold cross-validation accuracy for book attribution in 19th century literature.

6.1.2 Features

The feature sets used were FW, Con, Mod, and Com, as well as various combinations of these sets. Neither of the appraisal feature sets (App and Att) proved useful at all, nor did Con, Mod, or Com alone, so those results will not be shown.

6.1.3 Results

We ran 10-fold cross-validation tests using SMO (as described above) for classification of chapters for book, author, and author nationality (American or British). Results are shown in Figures 1, 2, and 3. Systemic features perform above baseline in all cases, though not as well as FW. However,

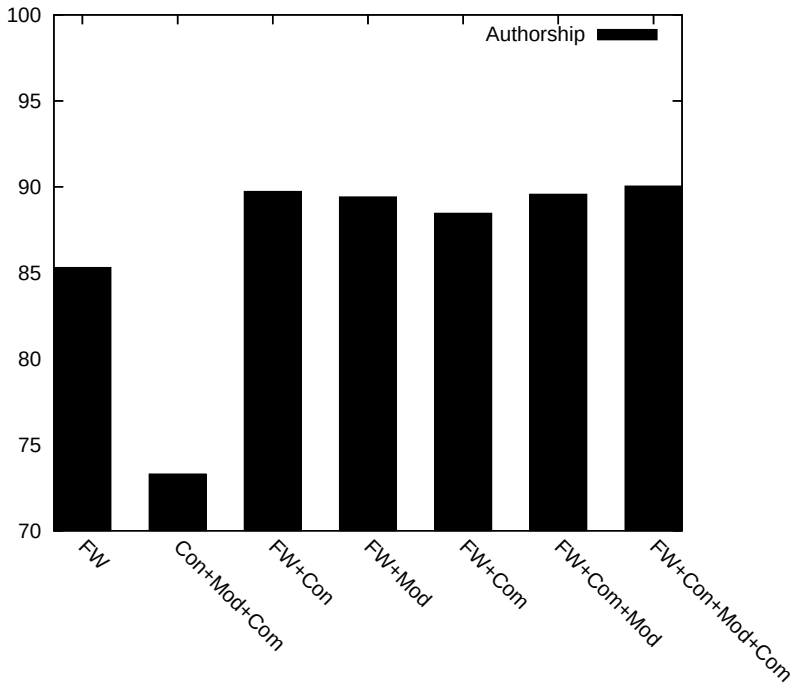

Figure 2: 10-fold cross-validation accuracy for authorship attribution in 19th century literature.

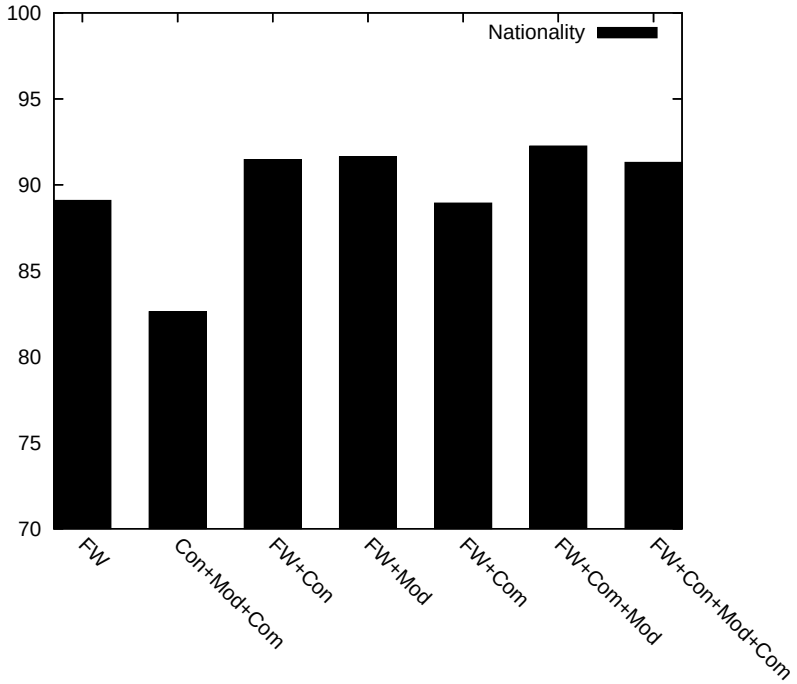

Figure 3: 10-fold cross-validation accuracy for nationality attribution in 19th century literature.

Table 2: The top ten features (by rank sum) for each of book, authorship, and nationality attribution.

Book Attribution	**Authorship Attribution**	**Nationality Attribution**
CONJ/Elaboration	CONJ/Enhancement	COMM/Assertive
upon	CONJ/Elaboration	MANIF/Explicit:ORIENT/Objective
CONJ/Enhancement	MODAL/Probability:ORIENT/Objective	COMM/Tentative
I	EXTENSION/Adversative	COMM/Presumptive
but	COMM/Presumptive	MODAL/Usuality:VAL/High
he	TYPE/Modulation:VAL/Median	MODUL/Readiness:VAL/High
she	MODAL/Probability:VAL/High	was
TYPE/Modulation:VAL/Median	SPATIOTEMPORAL/Complex	nobody
her	but	MODUL/Readiness:VAL/Low
with	COMM/Assertive	COMM/Admissive

in all cases, addition of systemic features improves classification accuracy. Authorship and book attribution were aided most by Con indicating useful variability in how authors structure and link information in their narratives, whereas nationality attribution was aided most by Com+Mod, indicating a possibly interesting difference between US and British authors in how events and propositions are assessed.

Note also that in this last case, we see that the addition of features may sometimes *reduce* classification accuracy. This is likely due to *overfitting*, where the presence of too many degrees of freedom in creating a model from training data leads the model to fit the training data too tightly, reducing its accuracy when applied to out-of-training test data. This is a well-known issue in machine learning (Schaffer, 1991); in this case it likely indicates again that conjunctions are not the strongest feature by which nationality should be classified.

To better understand which features contributed most to aid classification for each task, we considered the absolute weights assigned to them in the various linear models constructed by SMO for the highest-accuracy feature set in each task. For each task, features were ranked in each linear model (the multi-class problems use multiple linear models for classification) and the rank of each feature in each model were summed, producing an overall measure of the influence of the feature on classification for that task. The top ten features for each of the three tasks are shown in Table 2.

Examination of the features shows clear differences among the three tasks, in terms of which sorts of features were most significant for classification, from which we can draw some tentative conclusions. The three dominant categories of features for Book discrimination are: personal pronouns (four features), CONJUNCTION (two features), and prepositions (two features). The pronouns

likely indicate that narrative viewpoint and dialogue structure are important for distinguishing books[3], while the CONJUNCTION system and the prepositions 'upon' and 'with' likely indicate important variation in information organization and structure. In Author discrimination, we see four CONJUNCTION features, three MODALITY features, and two COMMENT features. Thus, while CONJUNCTION, hence information structure, is important both for distinguishing books and for distinguishing authors, we also see that different authors vary more in how they express MODALITY and COMMENTs, than do different books (by the same author). Finally, when considering the large-scale discrimination between American and British authors, we see MODALITY and COMMENT dominating, each with four features, indicating no significant difference in typical information structure between the two types of English, but definite variety in how MODALITY and COMMENTs are expressed. Further detailed studies on larger corpora will be needed to further elucidate these points.

6.2 Characterizing gender

We next examine the possibility of determining the gender (male or female) of literary characters based on their speeches; for this we considered characters from Shakespeare's plays. This extends previous results on classifying author gender (Argamon et al., 2003; de Vel et al., 2002) to examine the new question of whether a playwright can create recognizable distinctions between male and female characters.

6.2.1 The corpus

We constructed a corpus of characters' speeches from 38 Shakespearean plays, based on text from the Moby Shakespeare (Shakespeare, n.d.). A text file for each character in each play was constructed by concatenating all the character's speeches in the play; characters' genders were catalogued. To improve robustness of the results, all characters with less than 200 total words in the corpus were discarded. We further balanced the corpus for gender by keeping all 89 female characters (with at least 200 words) together with the 89 male characters with the most words each, discarding the rest. The composition of the corpus is summarized in Table 3.

[3]Thanks to David Hoover for this insight.

Table 3: Summary of the corpus of Shakespeare characters' speeches.

	Male		Female	
Play Name	**Num**	**Avg. Len**	**Num**	**Avg. Len**
All's Well That Ends Well	4	2537	4	1738.5
As You Like It	2	2362	3	2797
Cymbeline	5	2583	2	2734
Loves Labours Lost	1	2384	4	1000
Measure for Measure	3	3522	2	1698
Midsummer Nights Dream	0	n/a	3	1393.0
Much Ado About Nothing	1	2390	4	977
Pericles Prince of Tyre	1	4688	4	757
The Comedy of Errors	2	2346	4	904
The Merchant of Venice	2	2737	3	1914
The Merry Wives of Windsor	2	3104	3	1968
The Taming of the Shrew	1	3892	2	1080
The Tempest	1	4880	1	850
Troilus and Cressida	4	3262	1	2133
Twelfth Night	3	2384	3	2019
Two Gentlemen of Verona	1	3284	3	1376
Winter's Tale	3	3031	3	1581
The First part of King Henry IV	4	4155	2	342
The Second part of King Henry IV	2	4279	2	1138
The Life of King Henry V	1	8360	2	406
The First part of King Henry VI	1	1910	2	302
The Second part of King Henry VI	4	2492	2	2406
The Third part of King Henry VI	2	2802	3	708
The Life of King Henry VIII	2	3032	2	1672
The Life and Death of King John	3	3136	0	n/a
The Life and Death of Richard II	3	3790	2	829
The Life and Death of Richard III	2	4372	4	1524
Antony and Cleopatra	3	3692	2	2432
King Lear	3	3528	2	1086
Othello	3	5481	3	1551
Romeo and Juliet	4	2856	3	2397
The Life and Death of Julius Caesar	3	3848	1	720
The Tragedy of Coriolanus	4	3286	2	1360
The Tragedy of Hamlet	5	4618	1	1325
Timon of Athens	2	4113	0	n/a
Titus Andronicus	2	4048	5	1225

Table 4: Oppositions from the 20 highest-ranked features indicating each of Male and Female characters in the Shakespeare corpus. Features are ordered for easy reading, not by weight.

Condition	**Male**	**Female**
CONJUNCTION	Elaboration Enhancement	Extension
EXTENSION	Adversative	Additive
ENHANCEMENT	Causal/Conditional	Spatiotemporal Manner Matter
COMMENT	Admissive	Assertive Desiderative Evaluative Predictive Presumptive Tentative Validative
MODALITY TYPE	Modalization	Modulation

6.2.2 The features

The feature sets used were FW, Con, Mod, Com, App, and their various combinations.

6.2.3 Results

Classification accuracies under 10-fold cross-validation are shown in Figure 4. Function words give cross-validation accuracy of 68.5%, a small but noticeable effect (though not as high as previous results on male/female authorship classification (Koppel, Argamon, & Shimoni, 2003)). Considerably higher accuracy in this case is obtained from Con (74.7%) and slightly more yet by including more systemic features (both Con+Mod+Com and Con+Mod+Com+App give 75.8%). Note that Con+Mod+Com contains just 94 features as opposed to the 675 features in FW. Combining the two sets, however, gives an accuracy barely higher than that of FW by itself, indicating that FW is likely leading to overfitting for this problem.

We now consider what the most indicative features might say about the difference between male and female speech in Shakespeare's plays. Table 4 shows the oppositions in the 20 top male- and 20 top female-indicating features in the model learned for Con+Mod+Com. Several interesting differences between male and female characters are evident.

First we see that when using CONJUNCTION, female characters prefer Extension, whereas

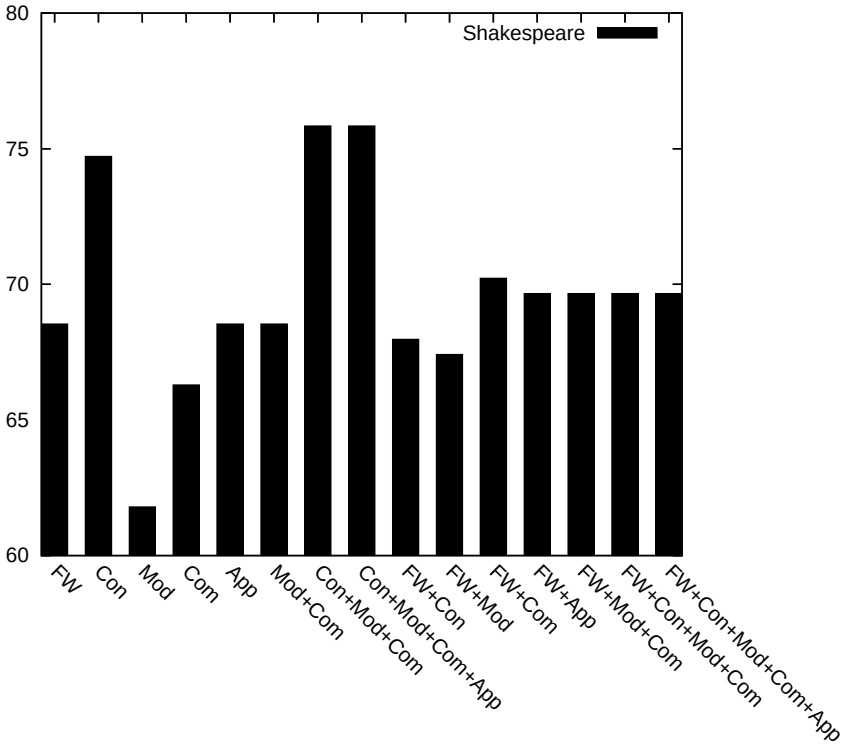

Figure 4: 10-fold cross-validation accuracy for discriminating male and female characters in Shakespeare.

male characters prefer the other two types. Generally speaking, this indicates that female speeches comprise more independent information units (linked by extension), whereas male speeches are more focused on a relatively smaller variety of information units. This may imply that Shakespeare's male characters have comparatively more soliloquies[4]. This notion is supported by measuring the number of individual speeches that are greater than four lines, where we see that the male characters in our corpus have a total of 2388 such 'long' speeches (26.8 on average), whereas the female characters have a total of 1027 long speeches (11.5 on average).

Next, going deeper within CONJUNCTION, we see that female characters prefer Additive EXTENSION, whereas male characters prefer Adversative EXTENSION. This says that when combining disparate information units, Shakespeare's males are more likely to "contrast and compare" than his females. We further note that within ENHANCEMENT, male characters prefer Causal and Conditional constructs more than female characters do (for whom, consequently, all other ENHANCEMENT types are indicative). Both of these facts argue that Shakespeare's male characters are more likely to construct complex logical arguments than his female characters. This hypothesis will need of course further confirmation from more detailed textual analysis.

Third, Admissive COMMENTs are the one type strongly preferred by male characters (and hence other COMMENT types are indicative of female characters). This might indicate that male characters are more concerned with issues of personal status (which an admission may affect) than female characters. This is quite plausible, though further analysis will be required.

Finally, within MODALITY, we see male characters preferring Modalization (typicality) and female characters preferring Modulation (readiness, obligation). This opposition has a straightforward (if stereotyped) interpretation that the male characters tend to construe the world in naturalistic terms (where impersonal notions of probability and usuality are in the fore), whereas the female characters construe things in a more intentional fashion (more concerned with obligation or readiness to do things).

While these features seem intuitively plausible as modes of characterizing male and female characters, a further analysis of male- and female-authored texts using these feature sets will be required to see whether Shakespeare's creation of gendered language corresponds to actual gender differences in language use extant in his time, and how it compares to that of other playwrights.

[4]Extended speeches on a single topic.

Table 5: Oppositions from the fifteen highest-ranked features indicating High and Low neuroticism in the stream-of-consciousness writing assignment. Features are ordered for easy reading, not by weight.

Condition	**High**	**Low**
ORIENTATION	Negative	Positive
ATTITUDE	Affect	Appreciation
APPRECIATION	Reaction-Quality	Reaction-Impact
APPRECIATION	Composition-Balance	Composition-Complexity
SOCIAL-ESTEEM	Tenacity	Normality
APPRECIATION/Valuation	Negative	Positive
APPRECIATION/Reaction-Quality	Negative	Positive
APPRECIATION/Reaction-Impact	Positive	Negative
APPRECIATION/Composition-Complexity	Positive	Negative
JUDGEMENT/Social-Sanction	Positive	Negative

6.3 Personality typing

We now turn back from considering fictional characters to considering real people, examining to what extent functional lexical features may usefully indicate an author's personality type. Specifically, we are concerned here with determining if an author has High or Low Neuroticism (roughly defined as the tendency to worry; see (Pervin & John, 2001)).

6.3.1 The corpus

The corpus used for this experiment was derived from student essays gathered by Prof. James Pennebaker at the University of Texas at Austin. As part of their course responsibilities, subjects (undergraduate students) wrote a stream-of-consciousness essay and an essay of deep self-analysis; these data sets (collected from 1997 and 2003) comprised 1157 and 1106 documents, respectively. Subjects were also given the NEO-FFI Five-Factor Personality Inventory (McCrae & P. T. Costa, 1996). Scores from the Neuroticism factor were used to define a binary classification task: Subjects with scores in the top third were classed as High, and those with scores in the bottom third classed as Low. The task is to use textual features to determine whether the author has High or Low neuroticism.

6.3.2 Features

The feature sets used were: FW, Con, Mod, Com, App, and their various combinations.

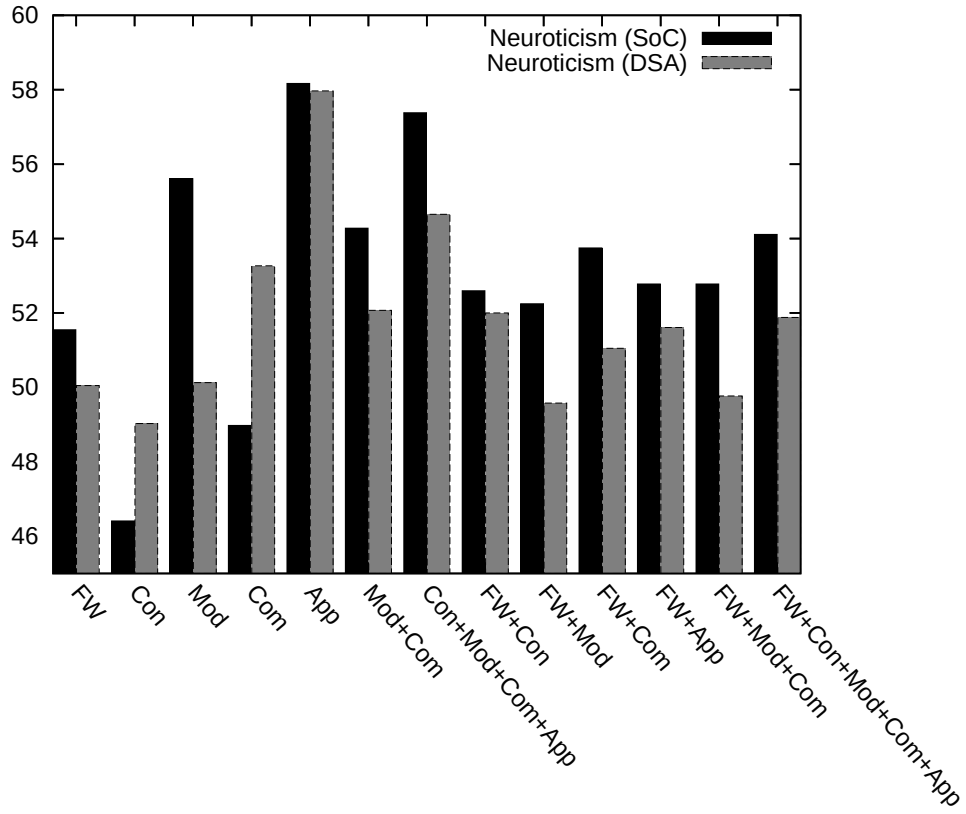

Figure 5: 10-fold cross-validation accuracy for discriminating High from Low neurotics in the stream-of-consciousness (SoC) and deep self-analysis (DSA) writing tasks.

Table 6: Oppositions from the fifteen highest-ranked features indicating High and Low neuroticism in the self-analysis writing assignment.

Condition	**High**	**Low**
ORIENTATION	Negative	Positive
ATTITUDE	Affect	Appreciation
JUDGEMENT	Social-Sanction	Social-Esteem
GRADUATION	Focus	Force
INTENSIFICATION	Maximization	High & Low
ATTITUDE/Appreciation	Negative	Positive
JUDGEMENT/Social-Esteem	Negative	Positive
APPRECIATION/CompositionBalance	Positive	Negative

6.3.3 Results

Accuracy results are shown in Figure 5; the task is clearly quite difficult as the effect of personality is weak (as previously noted (Pennebaker, Mehl, & Niederhoffer, 2003; Oberlander & Gill, 2004)). While the SoC texts were slightly more distinguishable, in both cases the most useful feature set for this task was Appraisal, with accuracies of 58.2% (SoC) and 58.0% (DSA). We anticipate that increasing the coverage of the Appraisal feature set to include also verbs and nouns will likely improve results.

The fact that Appraisal features gave the highest accuracy indicates (unsurprisingly) that a key difference between High and Low neurotics is in how they engage with and assess objects and people in their environment. A more detailed look at the specific features indicating either High or Low neuroticism can shed more light on the linguistic differences. All the oppositions found in the top fifteen features for High and Low neuroticism[5] are given in Tables 5 and 6.

First we consider the two oppositions that appear for both writing tasks[6]. Unsurprisingly, High neuroticism is associated with negative appraisal, whereas Low neuroticism is associated with positive appraisal. More interestingly, we see that the appraisal attitude expressed by High tends to is about Affect, reflecting a more self-directed focus on personal feelings, whereas that expressed by Low neurotics is about Appreciation, reflecting a more outer-directed focus that conceptualized appraisal as inherent attributes of external entities.

Most oppositions that appear for only one of the writing tasks reflect the general preference

[5]Fewer than fifteen oppositions appear, since a number of top-ranked features were unpaired.

[6]Note that none of the other oppositions are contradictory, allowing that these linguistic oppositions are consistent across different text types

of High neurotics for negative appraisal and Low neurotics for positive appraisal. However several oppositions give reversals of this general trend, to wit: Reaction-Impact, Composition-Complexity, Composition-Balance, and Social-Sanction. To understand this, note that these tend to be features generally preferred by Low Neurotics, hence generally avoided by High Neurotics (the one exception is Social-Sanction, in the self-analysis essays). It may therefore be that High Neurotics are more likely to use constructs they generally avoid when the feeling is Positive.

6.4 Sentiment analysis

Sentiment classification is the task of labelling a text as positive ('thumbs up') or negative ('thumbs down') based on the sentiment expressed by the author towards a target object (film, book, product, etc.). Important current applications include data and web mining, market research, and customer relationship management.

6.4.1 Corpus

To test the usefulness of adjectival appraisal groups for sentiment analysis, we evaluated the effectiveness of the above feature sets for movie review classification, using the publicly available collection of movie reviews constructed by Pang and Lee (2004). This standard testbed consists of 1000 positive and 1000 negative reviews, taken from the IMDb movie review archives[7]. Reviews with 'neutral' scores (such as three stars out of five) have been removed, giving a data set with only clearly positive and negative reviews.

6.4.2 Features

Since the only lexical taxonomy in this paper that is relevant to sentiment is Appraisal, the features used for sentiment analysis were Att and App (FW did not achieve appreciable accuracy). In addition, since many content-bearing words bear sentiment of various types, we also included a "Bag-of-Words" feature set (BoW), defined as the relative frequencies (as for FW) of all words in the corpus. Combinations of FW and BoW with both appraisal feature sets were also considered.

[7] See `http://www.cs.cornell.edu/people/pabo/movie-review-data/`

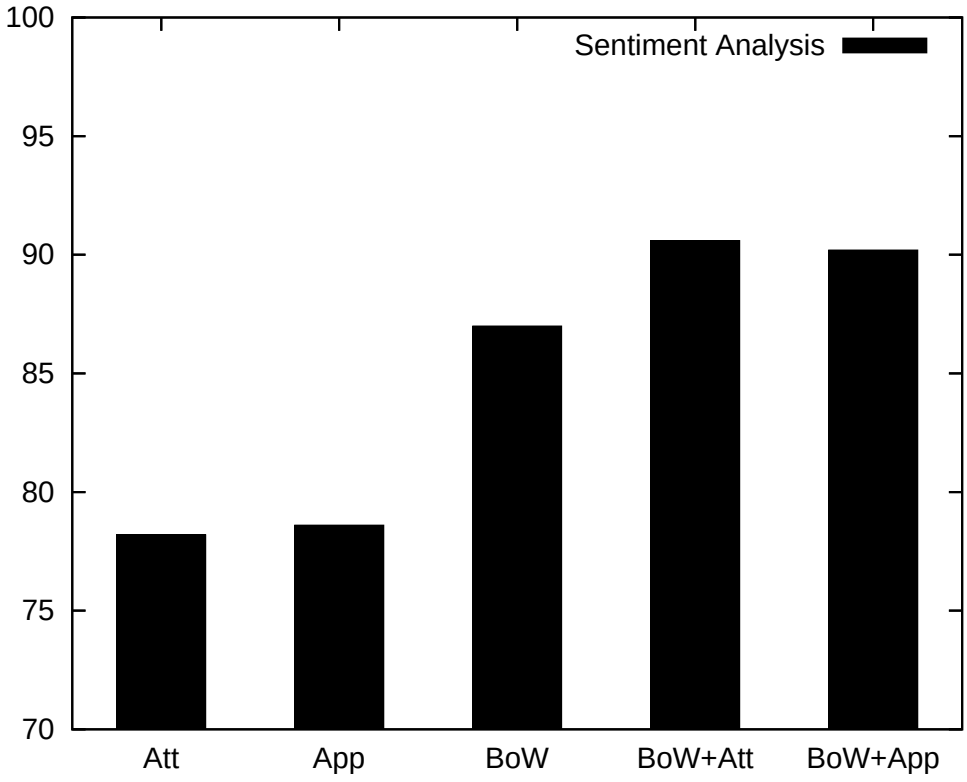

Figure 6: Movie review classification results for using SMO with default parameters and a linear kernel with various feature sets; accuracy is measured by 10-fold cross-validation.

6.4.3 Results

Figure 6 gives 10-fold cross-validation accuracy for different feature sets on sentiment analysis. The baseline of using just attitude-bearing adjectives is reasonably high, at 78.2% accuracy. This bears out our hypothesis that attitude-bearing adjectives specifically are a key feature in the expression of sentiment. Using attitude type and orientation of these terms yields a small improvement in accuracy to 78.2%.

Next, we note that all of the limited-coverage appraisal feature sets are outperformed by standard bag-of-words classification using all words (BoW), which attains 87.05% accuracy using SMO, competitive with Pang and Lee's (2004) recent result on this dataset, based on classifying texts after automatically extracting subjective passages from them. More significantly, we clearly improve on that result, attaining 90.6% accuracy, by combining appraisal features (Attitude Type) with the bag-of-words features (for coverage), demonstrating how appraisal analysis helps sentiment classification.

Of the most significant 30 features in the best model constructed, the vast majority (19) are drawn from subtypes of *appreciation*, with 6 (five negative and one positive) from *judgement* and two (one positive and one negative) from affect. Appreciation thus appears to be the most central type of attitude for sentiment analysis (at least for movie review classification). In addition, while some adjectival features in BoW are included (duplicating work done by Att), many BoW features are helping directly to fill known gaps in our current appraisal lexicon, including many nouns (e.g., 'mess', 'director', 'nothing', 'job', 'truth') and some verbs ('loved', 'wasted', 'delivered'), as well as other parts-of-speech.

From our perspective, most previous sentiment classification research has focused almost exclusively upon Orientation, with Attitude type addressed only indirectly, through the use of bag-of-words features. The only exceptions that we are aware of are recent works attempting to automatically determine the attitude type of a term. Kamps et al. (2002) used link-distances in WordNet to estimate three parameters, potency, activity, and evaluativity, based on Osgood et al.'s (1957) theory of semantic distances. One important exception is Taboada and Grieve's (2004) method of automatically determining top-level attitude types via application of of Turney's (2002) PMI method. They observed that different types of reviews contain different amounts of each attitude-

type, which our results confirm. Since the appraisal taxonomies used in this work are general purpose, and were not developed specifically for sentiment analysis or movie review classification, we expect appraisal group analysis to be highly portable to other related tasks.

6.5 Scientific prose

Finally, we ask whether different scientific fields have meaningfully distinctive language styles. To do so, we see if our functional lexical features can effectively classify peer-reviewed scientific articles from different fields, and if they can, whether the most indicative features give us any insight. (Also see our related study (Argamon, Dodick, & Chase, 2005) of experimental and historical science articles.)

6.5.1 Corpus

For this experiment we used a corpus comprising peer-reviewed articles from two geology and two paleontology journals (see Table 7):

Journal of Metamorphic Geology focuses on metamorphic studies[8], from the scale of individual crystals to that of lithospheric plates.

Journal of Geology includes research on the full range of geological principles including geophysics, geochemistry, sedimentology, geomorphology, petrology, plate tectonics, volcanology, structural geology, mineralogy, and planetary sciences.

Quaternary Research published research in diverse areas in the earth and biological sciences which examine the Quaternary period of the Earth's history (from roughly 1.6 million years ago to the present).

Paleontologica Electronica publishes papers in all branches of paleontology as well as related biological or paleontologically-related disciplines.

6.5.2 Features

The feature sets used were FW, Con, Mod, Com, App, and their various combinations.

[8]Metamorphism refers to changes in mineral assemblage and texture in rocks that have been subjected to temperatures and pressures different from those under which the rocks originally formed.

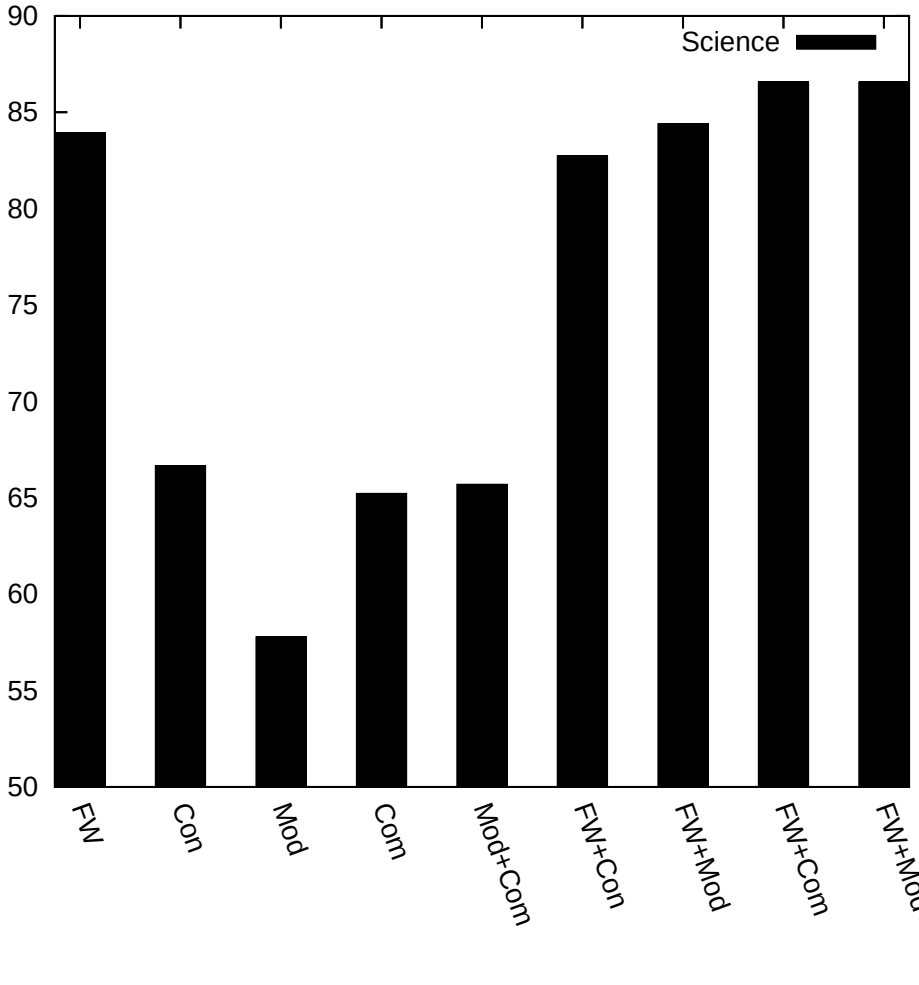

Figure 7: 10-fold cross-validation accuracies for SMO with various feature sets, on the corpus of science articles.

Table 7: Summary of the geology and paleontology journals used in the scientific literature corpus study, giving the number of articles from each journal in the corpus, and the average number of words per article.

Journal	**# Art.**	**Avg. Words**
Journal of Metamorphic Geology	108	5025
Journal of Geology	93	4891
Quaternary Research	113	2939
Paleontologia Electronica	111	4133

Table 8: Oppositions from the fifteen highest-ranked systemic features in Geology and Paleontology articles, from the model learned using FW+Com+Mod.

Condition	**Geology**	**Paleontology**
COMMENT	Predictive Admissive Evaluative Desiderative	Validative Assertive
MODALITY VALUE	Low	High
MODALITY MANIFESTATION	Implicit	Explicit
MODALITY MANIFESTATION/Implicit	VALUE/Median	VALUE/High
MODALITY MANIFESTATION/Explicit	VALUE/Low	VALUE/Median
MODALITY ORIENTATION/Subjective	VALUE/Low	VALUE/Median
MODALIZATION/Probability	VALUE/High	VALUE/Median

6.5.3 Results

10-fold cross-validation accuracy using SMO and the above feature sets are given in Figure 7. The baseline of function words alone gave 83.9% accuracy. Com and Mod+Com, when added to FW, gave a small but noticeably improved accuracy of 86.6%. The fact that Con tended to harm accuracy indicates that the overall textual structures used in these two scientific fields are similar, as we would have expected.

To understand some more about the relevant stylistic differences between geology and paleontology, we next consider oppositions among the top ten systemic features from FW+Com+Mod across the two article classes (Table 8). Six out of the eight COMMENT options are strongly indicative of one of the two classes; indeed as noted COMMENT provides the main discriminative power of systemic features for this problem. Some of these features have relatively clear interpretations, while others are more opaque. The preference of geologists for Predictive COMMENTs is likely due to the fact that they use experimental physical and chemical methods in their work more often than paleontologists, to elucidate mineral composition. Admissive COMMENTs in these articles

nearly always denote agreement with previous researchers (e.g., "therefore *we concur* that..."); perhaps Geologists are more likely to test previous hypotheses, since Paleontological samples are more unique. Evaluative COMMENTs are used in several contexts in the texts, tending in Paleontology to be just used to describe technique (e.g., "if *properly* prepared...") or reasoning (e.g., "they *inadvertently* confused this with..."), while in Geology they seem to be used for a larger variety of purposes, including terminology (e.g., "it is *properly* termed allanite") and properties of minerals (e.g., " fractured along *suitably* oriented cleavage planes"); more detailed study of this phenomenon will be required. Less clear are Assertive COMMENTs, which tend to be used for emphasis (e.g., "There is *surely* more to it..."), and Desiderative COMMENTs, expressing hope for or disappointment with research results (e.g., "records are *unfortunately* much more fragmented..."); the reason for these differences between Geology and Paleontology is as yet unclear.

In MODALITY, the main difference is between Geology's general preference for expressing Low VALUE vs. Paleontology's preference for High VALUE (except in Probability). This general tendency may reflect the fact that most Geological conclusions are relatively sure (and so need no modal qualification), but those that are not sure are exceptional, and so are modally qualified, whereas in Paleontology, due to the paucity of specimens of any given type, few assertions are sure, so those that are will be more likely marked by MODALITY. The one exception here is Probability, where Geology prefers High VALUE and Paleontology Median (with no appreciable difference for Low VALUE). We hypothesize that Median VALUEs are more likely to express general disclaimers (consider 'probably' or 'usually') than either High (which express certainty) or Low (which express clear uncertainty); for the reasons just mentioned, we expect Paleontology articles to use more such hedges than Geology.

7 Conclusions

This paper has presented a novel set of linguistically motivated *functional lexical* features for stylistic text classification which have been shown to increase classification accuracy over the function word baseline for a variety of stylistic classification tasks, in some cases quite significantly. More significantly, our results show how different kinds of features are needed for different kinds of stylistic text classification, and that addition of irrelevant features often reduces performance (that

is, overfitting is difficult to avoid). Indeed, we have shown how the organization of these features in systemic taxonomies enables us to gain fairly detailed insights into the linguistic choices inherent in different text styles. This is because siblings in a taxonomy represent functional alternatives in the language that can be meaningfully compared. Thus one potentially important use of these methods is for *computational sociolinguistics*, extending the study of systematic linguistic variation from the small-scale studies typical of traditional sociolinguistics (Trudgill, 2001; Labov, 1973) to larger-scale studies examining lexical and grammatical variation related to geography, education, societal status, and other factors.

There are a number of directions we are currently pursuing to improve the methods presented in this paper. The most immediate is extending the coverage of the functional lexical taxonomies, both increasing their completeness, and adding new functional taxonomies. We are also currently developing methods for efficiently parsing phrases (such as adjectival groups) based on systemic functional principles; we recently presented early results (Whitelaw, Garg, & Argamon, 2005). We expect this to improve analysis by both enabling us to extract new functional attributes in a text, and also correcting inaccuracies in our current purely lexical approach (since a group may have different functional attributes than its head — compare "good" with "not good"). Such shallow parsing will also help reduce ambiguity; for example "even" on its own may be an adjective referring to APPRECIATION/QualityBalance, whereas in a phrase "not even nice" it is clearly seen to be a modifier. More generally, recent work in such areas as Rhetorical Structure Theory (Moore & Pollack, 1992; Marcu, 1997, 1999) collocation analysis (Wiebe, McKeever, & Bruce, 1998; Wiebe, Wilson, & Bell, 2001), and coreference resolution (Ng & Cardie, 2002; Cristea, Marcu, Ide, & Tablan, 1999; Schauer & Hahn, 2001) might also be applied to improve and extend extraction of functionally relevant stylistic features.

Acknowledgments

Many thanks to Ophir Frieder and Wai Gen Yee for their careful readings of this paper and many helpful comments. Also thanks to Jeff Dodick for interesting and productive discussions about the nature of the various earth sciences.

References

Androutsopoulos, I., Koutsias, J., Chandrinos, K., Paliouras, G., & Spyropoulos, C. (2000). An evaluation of naive bayesian anti-spam filtering. In Proc. of the workshop on machine learning in the new information age.

Argamon, S., Dodick, J., & Chase, P. (2005, July). The languages of science: A corpus-based study of experimental and historical science articles. In Proc. 26th annual meeting of the cognitive science society.

Argamon, S., Koppel, M., Fine, J., & Shimony, A. R. (2003). Gender, genre, and writing style in formal written texts. Text, 23(3).

Argamon, S., & Levitan, S. (2005, June). Measuring the usefulness of function words for authorship attribution. In Proceedings of the 2005 ACH/ALLC conference. Victoria, BC, Canada.

Baayen, R. H., Halteren, H. van, & Tweedie, F. (1996). Outside the cave of shadows: Using syntactic annotation to enhance authorship attribution. Literary and Linguistic Computing, 7, 91-109.

Belkin, N. J. (1993). Interaction with texts: Information retrieval as information-seeking behavior. In Information retrieval (p. 55-66).

Berry, M. J., & Linoff, G. (1997). Data mining techniques: For marketing, sales, and customer support. New York, NY, USA: John Wiley & Sons, Inc.

Burrows, J. F. (1987). Computation into criticism: A study of jane austen's novels and an experiment in method. Oxford: Clarendon Press.

Chaski, C. E. (1999). Linguistic authentication and reliability. In National conference on science and the law. San Diego, California: National Institute of Justice.

Chen, S. Y., Magoulas, G. D., & Dimakopoulos, D. (2005). A flexible interface design for web directories to accommodate different cognitive styles. JASIST, 56(1), 70-83.

Cristea, D., Marcu, D., Ide, N., & Tablan, V. (1999). Discourse structure and co-reference: An empirical study. In D. Cristea, N. Ide, & D. Marcu (Eds.), The relation of discourse/dialogue structure and reference (pp. 46–53). New Brunswick, New Jersey: Association for Computational Linguistics.

de Vel, O., Corney, M., Anderson, A., & G.Mohay. (2002, August). Language and gender author

cohort analysis of e-mail for computer forensics. In Proc. digital forensic research workshop. Syracuse, NY.

Finn, A., Kushmerick, N., & Smyth, B. (2002). Genre classification and domain transfer for information filtering. In F. Crestani, M. Girolami, & C. J. van Rijsbergen (Eds.), Proceedings of ECIR-02, 24th european colloquium on information retrieval research. Glasgow, UK: Springer Verlag, Heidelberg, DE.

Fritch, J. W., & Cromwell, R. L. (2001). Evaluating internet resources: identity, affiliation, and cognitive authority in a networked world. J. Am. Soc. Inf. Sci. Technol., 52(6), 498–507.

Grossman, D., & Frieder, O. (1998). Information retrieval: Algorithms and heuristics. Kluwer Academic Publishers.

Halliday, M. A. K. (1994). Introduction to functional grammar (Second ed.). Edward Arnold.

Halliday, M. A. K., & Hasan, R. (1976). Cohesion in english. Longman.

Hammer, J., Garcia-Molina, H., Cho, J., Crespo, A., & Aranha, R. (1997). Extracting semistructured information from the web. In Proceedings of the workshop on management for semistructured data.

Holmes, D. I. (1998). The evolution of stylometry in humanities scholarship. Literary and Linguistic Computing, 13(3), 111–117.

Hoover, D. (2002). Frequent word sequences and statistical stylistics. Literary and Linguistic Computing, 17, 157-180.

Kamps, J., Marx, M., Mokken, R. J., & Rijke, M. de. (2002). Words with attitude. In Proc. 1st international conference on global WordNet. Mysore, India.

Karlgren, J. (2000). Stylistic experiments for information retrieval. Unpublished doctoral dissertation, SICS.

Karlgren, J., & Cutting, D. (1994). Recognizing text genres with simple metrics using discriminant analysis. In Proceedings of the 15th. international conference on computational linguistics (COLING 94) (Vol. II, pp. 1071 – 1075). Kyoto, Japan.

Kehagias, A., Petridis, V., Kaburlasos, V., & Fragkou, P. (2003). A comparison of word- and sense-based text categorization using several classification algorithms. Journal of Intelligent Information Systems, 21(3), 227–247.

Kessler, B., Nunberg, G., & Schütze, H. (1997). Automatic detection of text genre. In P. R. Cohen

& W. Wahlster (Eds.), Proceedings of the thirty-fifth annual meeting of the Association for Computational Linguistics and eighth conference of the European chapter of the Association for Computational Linguistics (pp. 32–38). Somerset, New Jersey: Association for Computational Linguistics.

Kjell, B., Woods, W. A., & Frieder, O. (1994, January). Discrimination of authorship using visualization. Information Processing and Management, 30(1), 141-150.

Koppel, M., Akiva, N., & Dagan, I. (2003). A corpus-independent feature set for style-based text categorization. In Workshop on computational approaches to style analysis and synthesis, 18th international joint conference on artificial intelligence. Acapulco.

Koppel, M., Argamon, S., & Shimoni, A. R. (2003). Automatically categorizing written texts by author gender. Literary and Linguistic Computing, 17(4).

Kushmerick, N. (1999). Learning to remove internet advertisement. In O. Etzioni, J. P. Müller, & J. M. Bradshaw (Eds.), Proceedings of the third international conference on autonomous agents (agents'99) (pp. 175–181). Seattle, WA, USA: ACM Press.

Labov, W. (1973). Sociolinguistic patterns. University of Pennsylvania Press.

Marcu, D. (1997). The rhetorical parsing of natural language texts. In Meeting of the association for computational linguistics (p. 96-103).

Marcu, D. (1999). A decision-based approach to rhetorical parsing. In Proc. of ACL'99 (pp. 365–372).

Martin, J. R., & White, P. R. R. (2005). The language of evaluation: Appraisal in english. London: Palgrave. ((`http://grammatics.com/appraisal/`))

Matthews, R. A. J., & Merriam, T. V. N. (1997). Distinguishing literary styles using neural networks. In Handbook of neural computation (chap. 8). IOP publishing and Oxford University Press.

Matthiessen, C. (1995). Lexico-grammatical cartography: English systems. International Language Sciences Publishers.

Matthiessen, C., & Bateman, J. A. (1991). Text generation and systemic-functional linguistics: experiences from english and japanese. London and New York: Frances Pinter Publishers and St. Martin's Press.

McCrae, R. R., & P. T. Costa, J. (1996). Toward a new generation of personality theories:

Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), The five-factor model of personality: Theoretical perspectives (pp. 51–87). New York: Guilford.

McEnery, A., & Oakes, M. (2000). Authorship studies/textual statistics. In Handbook of natural language processing. Marcel Dekker.

McKinney, V., Yoon, K., & Zahedi, F. M. (2002). The measurement of web-customer satisfaction: An expectation and disconfirmation approach. Info. Sys. Research, 13(3), 296–315.

McMenamin, G. (2002). Forensic linguistics: Advances in forensic stylistics. CRC Press.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). Wordnet: An on-line lexical database. International Journal of Lexicography, 3(4), 235-312.

Moore, J. D., & Pollack, M. E. (1992). A problem for RST: The need for multi-level discourse analysis. Computational Linguistics, 18(4), 537–544.

Mosteller, F., & Wallace, D. L. (1964). Inference and disputed authorship: The federalist. Massachusetts: Addison-Wesley.

Ng, V., & Cardie, C. (2002). Improving machine learning approaches to coreference resolution. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (pp. 104–111).

Oberlander, J., & Gill, A. (2004). Individual differences and implicit language: Personality, parts-of-speech and pervasiveness. In Proceedings of the 26th annual conference of the cognitive science society (pp. 1035–1040). Chicago, 2004.

Osgood, C. E., Succi, G. J., & Tannenbaum, P. H. (1957). The measurement of meaning. University of Illinois.

Pang, B., & Lee, L. (2004, July). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proc. 42nd ACL (pp. 271–278). Barcelona, Spain.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. In Proceedings of the 2002 conference on empirical methods in natural language processing (emnlp).

Patrick, J. (2004). The scamseek project: Text mining for finanical scams on the internet. In S. Simoff & G. Williams (Eds.), Proc. 3rd australasian data mining conf (p. 33-38). Carins.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. Annual Review of Psychology, 54, 547–577.

Pervin, L. A., & John, O. P. (Eds.). (2001). Handbook of personality theory and research (Second ed.). Guilford.

Platt, J. (1998). Fast training of support vector machines using sequential minimal optimization. In Advances in kernel methods – support vector learning. MIT Press.

Ponte, J. M., & Croft, W. B. (1998). A language modeling approach to information retrieval. In Research and development in information retrieval (p. 275-281).

Salton, G., & McGill, M. (1983). Introduction to modern information retrieval. McGraw-Hill.

Schaffer, C. (1991). Overfitting avoidance as bias. In IJCAI-91 workshop on evaluating and changing representation in machine learning. Sydney.

Schauer, H., & Hahn, U. (2001). Anaphoric cues for coherence relations. In G. Angelova, K. Bontcheva, R. Mitkov, N. Nicolov, & N. Nikolov (Eds.), Proceedings of the Euroconference Recent Advances in Natural Language Processing (RANLP-2001) (pp. 228–234). Tzigov, Bulgaria.

Sebastiani, F. (2002). Machine learning in automated text categorization. ACM Computing Surveys, 34(1), 1–47.

Shakespeare, W. (n.d.). The complete moby shakespeare. (http://www-tech.mit.edu/Shakespeare/)

Stamatatos, E., Fakotakis, N., & Kokkinakis, G. K. (2000). Automatic text categorization in terms of genre, author. Computational Linguistics, 26(4), 471-495.

Taboada, M., & Grieve, J. (2004). Analyzing appraisal automatically. In Aaai spring symposium on exploring attitude and affect in text. AAAI.

Tang, R., Ng, K. B., Strzalkowski, T., & Kantor, P. B. (2003). Toward machine understanding of information quality. In Proceedings of 2003 annual meeting of american society for information science and technology (Vol. 40, p. 213-220).

Teich, E. (1995). A proposal for dependency in systemic functional grammar – metasemiosis in computational systemic functional linguistics. Unpublished doctoral dissertation, University of the Saarland and GMD/IPSI, Darmstadt.

Trudgill, P. (2001). Sociolinguistics: An introduction to language and society (Fourth ed.). Penguin Books.

Turney, P., & Littman, M. (2002). Unsupervised learning of semantic orientation from a hundred-billion-word corpus (Tech. Rep. No. ERB-1094; NRC #44929). National Research

Council Canada.

Turney, P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In Proceedings 40th annual meeting of the acl (acl'02) (pp. 417–424). Philadelphia, Pennsylvania.

Tweedie, F., Singh, S., & Holmes, D. (1996). Neural network applications in stylometry: The Federalist Papers. Computers and the Humanities, 30(1), 1–10.

Vel, O. de. (2000). Mining e-mail authorship. In Workshop on text mining, ACM international conference on knowledge discovery and data mining. Boston, MA.

Whitelaw, C., Garg, N., & Argamon, S. (2005, May). Using appraisal taxonomies for sentiment analysis. In Proc. second midwest computational linguistic colloquium (MCLC 2005). Columbus, Ohio.

Wiebe, J., McKeever, K., & Bruce, R. (1998). Mapping collocational properties into machine learning features. In Proc. 6th workshop on very large corpora (pp. 225–233).

Wiebe, J., Wilson, T., & Bell, M. (2001). Identifying collocations for recognizing opinions. In Proceedings of ACL/EACL 2001 workshop on collocation.

Witten, I. H., & Frank, E. (2000). Data mining: Practical machine learning tools with java implementations. San Francisco: Morgan Kaufmann.

Yule, G. U. (1944). Statistical study of literary vocabulary. Cambridge University Press.

A Functional Lexical Taxonomies

This appendix describes in more detail the four types of functional lexical features used in the paper, with their associated taxonomies of attribute values.

A.1 Conjunction

Words and phrases that conjoin clauses (such as 'and', 'while', and 'in other words') are organized in SFG in the CONJUNCTION system network. Types of CONJUNCTION serve to link a clause with its textual context, by denoting how the given clause expands on some aspect of its preceding context (Matthiessen, 1995, p. 519–528). Similar systems also operate at the lower levels of noun and verbal groups, 'overloading' the same lexical resources while denoting similar relationships, e.g., *and* usually means "additive extension". The three top-level options of CONJUNCTION are Elaboration, Extension, and Enhancement:

- Elaboration: Deepening the content in its context by exemplification or refocusing.
- Extension: Adding new related information, perhaps contrasting with the current information.
- Enhancement: Qualifying the context by circumstance or logical connection.

A more detailed picture of the system network, with examples of lexical items corresponding to each option, is given in Figure 8.

Different patterns of CONJUNCTION usage lead to markedly different textual styles. Frequent use of Extension can give a text with high information density which can give a 'panoramic' effect of touring through a conceptual landscape, but if done poorly may overwhelm and lose a reader in too many facts. On the other hand, Elaboration can be used to good effect to create textual coherence around a single focused storyline. We note too that many of the standard function words traditionally used in computational stylistic studies are types of CONJUNCTION, which further argues for this system's importance for stylistic text analysis.

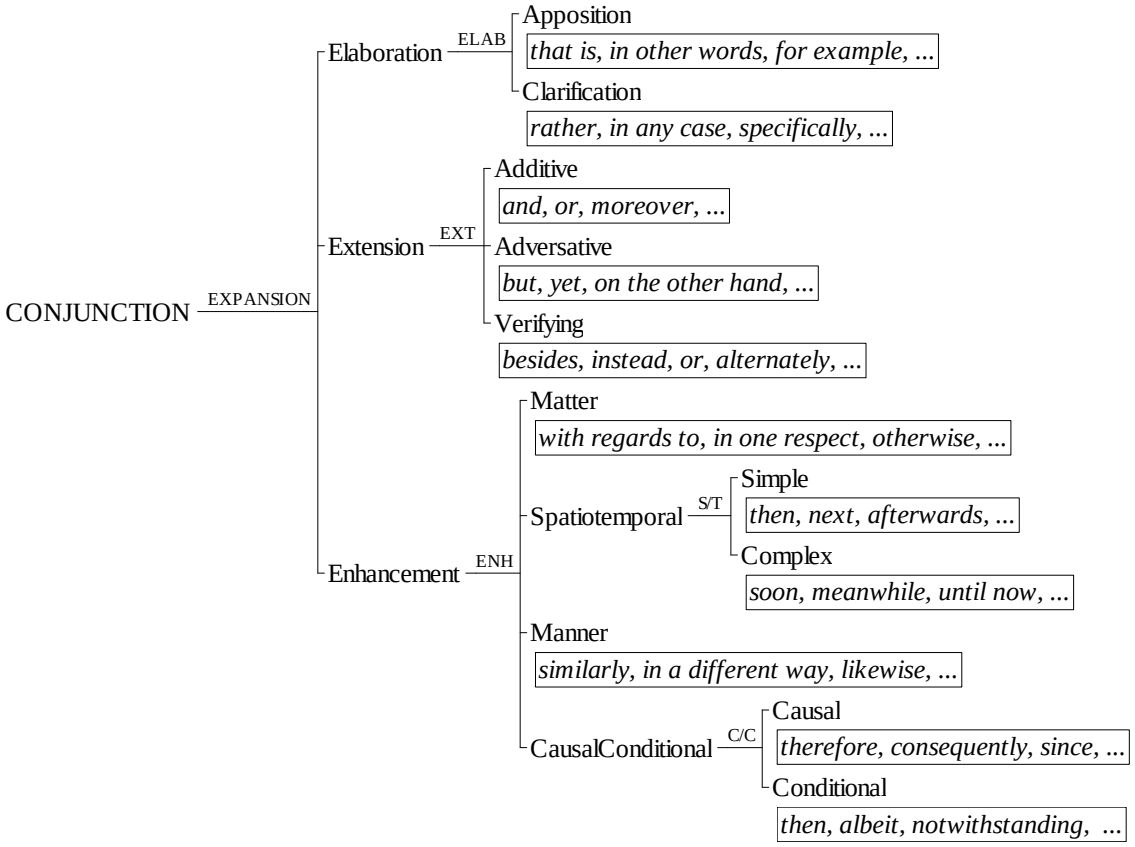

Figure 8: The CONJUNCTION system (Matthiessen, 1995). Options here are disjunctive; examples of lexical realizations for the leaves are given in italics.

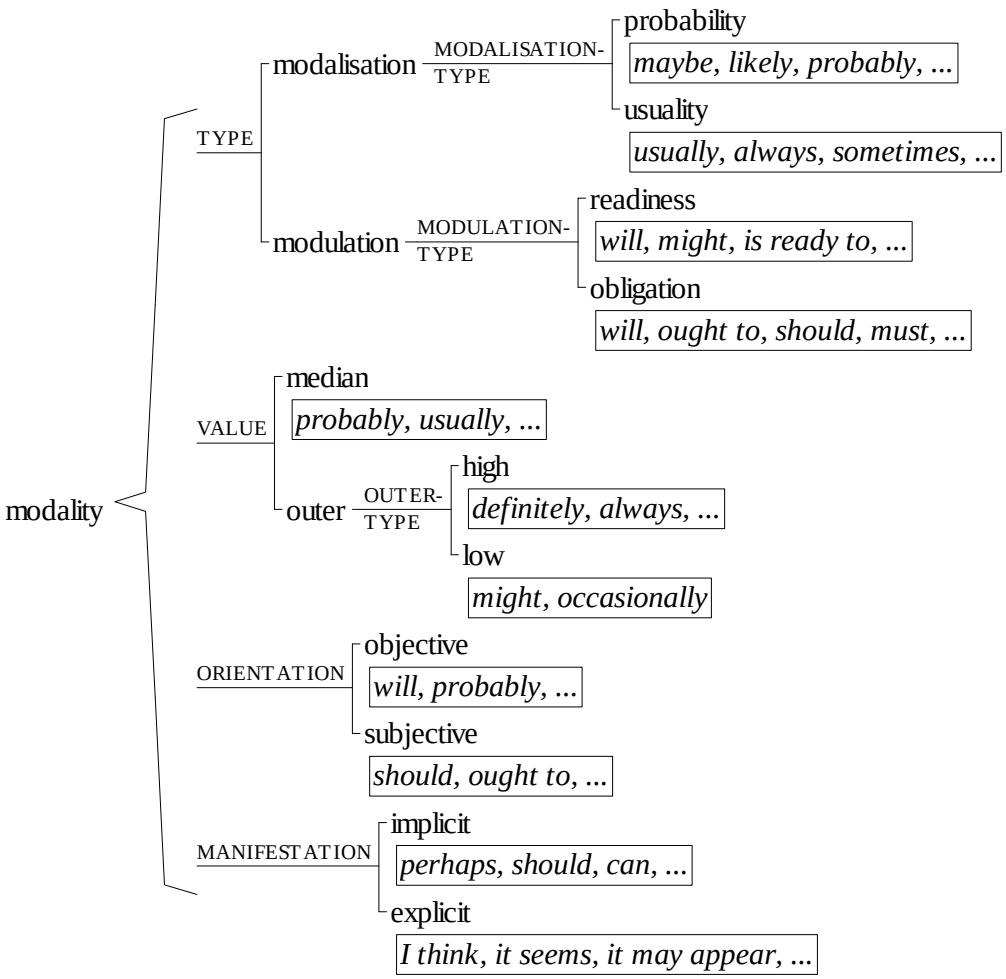

Figure 9: The MODALITY system networks (Matthiessen, 1995), notation as above.

			Type:Modalization		Type:Modulation	
			Probability	Usuality	Readiness	Obligation
Objective	Explicit	Median	*is likely*	*is frequent*	—	*is preferable*
Objective	Explicit	High	*is undeniable*	—	—	*is required*
Objective	Explicit	Low	*is possible*	*is infrequent*	—	*is permitted*
Objective	Implicit	Median	*probably*	*usually*	*eager to*	*ought to*
Objective	Implicit	High	*certainly*	*always*	*decided to*	*obliged to*
Objective	Implicit	Low	*maybe*	*seldom*	*allowed to*	*able to*
Subjective	Explicit	Median	*we believe*	—	*we prefer*	—
Subjective	Explicit	High	*we know*	—	—	*we require*
Subjective	Explicit	Low	*we suspect*	—	—	*we permit*
Subjective	Implicit	Median	*will*	*will*	*would rather*	*should*
Subjective	Implicit	High	*must*	*must*	*must, has to*	*ought to*
Subjective	Implicit	Low	*can, may*	*can, may*	*can, will*	*can, could*

Figure 10: Examples of indicator features for various combinations of MODALITY options. Note that not all combinations are realized in the language; note also the ambiguity of some of the indicators.

A.2 Modality

The system of MODALITY enables writers to qualify events or entities in the text according to their likelihood, typicality, or necessity. Syntactically, MODALITY may be realized in a text through a modal verb (e.g., 'can', 'might', 'should', 'must'), an adverbial adjunct (e.g., 'probably', 'preferably'), or use of a projective clause (e.g., "I think that...", "It is necessary that..."). Each expression of MODALITY has four attributes, corresponding to simultaneous choice of options within four system networks. These networks, with their top-level options, are as follows (the complete picture as we have implemented is shown in Figure 9):

- Type: What kind of modality is being expressed?
  - Modalization: How 'typical' is it?
  - Modulation: How 'necessary' is it?
- Value: What degree of the relevant modality scale is being averred?
  - Median: The 'normal' amount.
  - Outer: An extreme (either high or low) amount.
- Orientation: Relation of the modality expressed to the speaker/writer.
  - Objective: Modality expressed irrespective of the speaker/writer.
  - Subjective: Modality expressed relative to the speaker/writer.
- Manifestation: How is the modal assessment related to the event being assessed?
  - Implicit: Modality realized 'in-line' by an adjunct or modal auxiliary.
  - Explicit: Modality realized by a projective verb, with the nested clause being assessed.

Any given expression of MODALITY will choose options in parallel from these four networks, though some combinations are rare or non-existent. Figure 10 gives examples of lexical items for each possible combination of attributes.

A.3 Comment

The system of COMMENT provides a resource for the writer to 'comment' on the status of a message with respect to textual and interactive context in a discourse. Comments are usually realized as adjuncts in a clause and may appear initially, medially, or finally. Matthiessen (1995), following Halliday (1994), lists eight COMMENT options, as follows:

- Admissive: Message is an admission (e.g., "Frankly...")
- Assertive: Emphasis of reliability (e.g., "Certainly...")
- Desiderative: Desirability of the content (e.g., "Unfortunately...")
- Evaluative: Judgment of the actors involved (e.g., "Sensibly...")
- Predictive: Coherence with predictions (e.g., "As expected...")
- Presumptive: Dependence on other assumptions (e.g., "I suppose that...")
- Tentative: Assessing the message as tentative (e.g., "Tentatively...")
- Validative: Assessing scope of validity (e.g., "In general...")

A.4 Appraisal

Appraisal denotes how language is used to adopt or express an attitude of some kind towards some target. For example, in "I found the movie quite monotonous", the speaker adopts a negative *Attitude* ("monotonous") towards "the movie" (the *appraised object*). Note that attitudes come in different types; for example, 'monotonous' describes an inherent quality of the appraised object, while 'loathed' would describe an emotional reaction of the writer. The overall type and orientation of appraisal expressed in the text about an object gives a picture of how the writer wishes the reader to view it (modulo sarcasm, of course). To date, we have developed a lexicon for appraisal adjectives as well as relevant modifiers (such as 'very' or 'sort of'); we are currently developing a shallow parser that will be able to extract adjectival appraisal groups as well as identify the appraiser and the appraised object.

Following Martin and White (2005), we define five appraisal attributes: Attitude, Orientation, Force, Focus, and Polarity[9]:

[9]Note we use the term 'Polarity', in its SFG sense, to denote the grammatical notion of "explicit negation of a

Attitude gives the type of appraisal being expressed as either *affect*, *appreciation*, or *judgment*. Affect refers to a personal emotional state (e.g., 'happy', 'angry'), and is the most explicitly subjective type of appraisal. The other two options express evaluation of external entities, differentiating between evaluation of intrinsic *appreciation* of object properties (e.g., 'slender', 'ugly') and social *judgment* (e.g., 'heroic', 'idiotic'). Figure 11 gives a more detailed view of the various options in Attitude, together with illustrative adjectives of each type. In general, attitude may be expressed through nouns (e.g., 'triumph', 'catastrophe') and verbs (e.g., 'love', 'hate'), as well as adjectives; we are currently working on expanding our lexicon to include nouns and verbs as well.

Orientation is whether the appraisal is *positive* or *negative* (often termed 'sentiment').

Force denotes the intensity of the appraisal being expressed; for example, "good" will have neutral Force, while "great" will have high Force, and "the best" will have maximal Force.

Focus is another aspect of the graduation of appraisal, referring to the 'prototypicality' of the appraisal being expressed; for example the modifier "truly..." is a Focus sharpener, while "sort of..." is a Focus softener.

Polarity of an appraisal is *marked* if it is scoped in a polarity marker (such as 'not'), or *unmarked* otherwise. Other attributes of appraisal are, of course, affected by negation; for example, "not good" expresses a different sentiment from "good".

Appraisal adjectives take on attribute values from all five appraisal attributes as described above. Appraisal modifiers, on the other hand, have values for just the latter four attributes, as Attitude type cannot be modified.

A value for each appraisal attribute is stored for each appraisal adjective; for example, the

quality or assertion within the scope of the particle 'not' or the equivalent", although this term has sometimes been used to mean what we refer to as 'Orientation'.

lexical entry for ‘beautiful’ reads:

‘beautiful’	
Attitude:	appreciation/reaction-quality
Orientation:	positive
Force:	neutral
Focus:	neutral
Polarity:	unmarked

Modifiers, mostly adverbs, give transformations for one or more appraisal attributes, for example:

‘very’	
Force:	increase

or polarity modification:

‘not’	
Orientation:	negate
Force:	reverse
Polarity:	marked

Modifiers can specify effects on multiple appraisal attributes at once; e.g., ‘really’ functions both as an intensifier of force and a sharpener of focus.

The experiments reported in this paper only consider the Attitude and Orientation attributes of appraisal adjectives; elsewhere, we have presented some early results using shallow parsing of adjectival groups (Whitelaw et al., 2005).

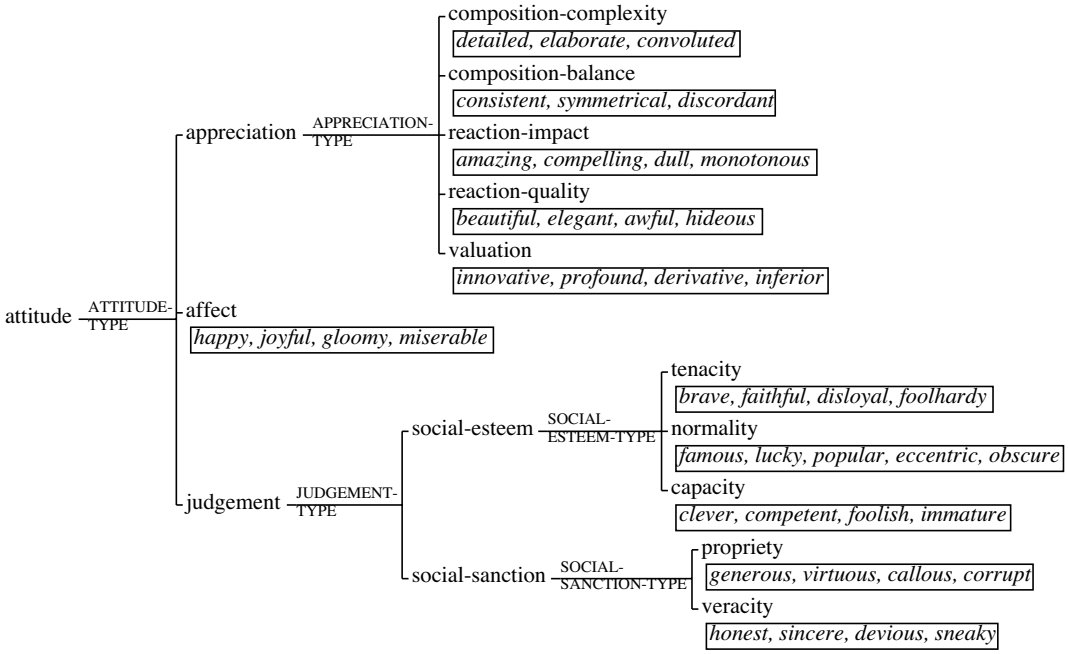

Figure 11: Options in the Attitude network, with examples of appraisal adjectives from our lexicon.