Lexical Predictors of Personality Type[*]

Shlomo Argamon[1], **Sushant Dhawle**[1], **Moshe Koppel**[2], **James W. Pennebaker**[3]

1. Dept. of Computer Science, Illinois Institute of Technology, Chicago, IL 60616
2. Dept. of Computer Science, Bar-Ilan University, Ramat Gan 52900, Israel
3. Dept. of Psychology, The University of Texas, Austin, TX 78712

Abstract

We are currently pursuing methods for "author profiling" in which various aspects of the author's identity might be identified from a text, without necessarily having a corpus of documents from the same individual. A key component of such an identity profile is personality; this paper addresses distinguishing high from low neuroticism and extraversion in authors of informal text. We consider four different sets of lexical features for this task: a standard function word list, conjunctive phrases, modality indicators, and appraisal adjectives and modifiers. SMO, a support vector machine learner, was used to learn linear separators for the high and low classes in each of the two tasks. We find that appraisal use is the best predictor for neuroticism, and that function words work best for extraversion. Further, examination of the specifically most important features yields insight into how neuroticism and extraversion differentially affect language use.

1 Introduction

The ways individuals use words can reflect basic psychological processes, including clues to their thoughts, feelings, perceptions, and personality. Using recent developments in machine learning and language psychology, we seek reliable linguistic markers of social and psychological states. Examples of the questions we hope to eventually answer reliably include: Is the author male or female? How old is the author? To what degree does the author display signs of depression, high self-esteem, or other personality characteristics? Can we determine if the author is writing or speaking during times of high stress, grief, or in the throes of love? The system we are constructing is able to generate language-based predictive models for answering questions such as these based on text samples. By eventually using texts in multiple languages and from multiple (Western) cultures, we hope to gain insight into the complex relations between language, psychological processes, and culture.

There is great potential in this area for important practical applications both in the area of literary research and in the area of forensics. Consider the typical situation in which we are faced with an anonymous document that we wish to attribute to one of a very large class of suspected authors. The usual techniques will not work except in the unusual circumstance that we have significant quantities of material authored by each of the suspects. What can we do under ordinary circumstances where such material is unavailable? The solution is to use the copious material we do have available, authored by males and females of various ages, linguistic histories and personality types, to "profile" the suspect. Thus, even if a suspect can't be uniquely identified, the list of viable suspects can be reduced to those that satisfy a given profile.

This paper focuses on determining "personality type" of the author from casual written text. We focus on two key dimensions of personality: Neuroticism (roughly: tendency

---

to worry), and Extraversion (roughly: preference for the company of others). By automatically analyzing the style of written texts to determine personality type using machine learning, our work integrates two existing strands of research in *language psychology* and *computational stylistics*.

1.1 Language psychology

A central assumption of language psychology is that the words people use reflect who they are. Beginning in the 1950s, a small group of researchers in psychology and medicine discovered that the ways people spoke were related to their physical and mental health problems (e.g., [6, 7]). With increasing computer advances, strides were made in attempting to capture psychological themes or people's underlying emotional states that might be reflected in the words they used (e.g., [19, 29, 25]).

In recent years, function words (which have been shown to be processed in the brain differently from nouns and regular verbs) have been shown to mirror people's psychological states. When depressed or in an emotionally vulnerable situation, individuals exhibit increases in pronouns (especially first person singular), drops in articles, and increases in their use of present tense auxiliary verbs [22]. When facing a collective or shared upheaval, such as September 11, a local disaster, or learning about the death of a famous person, people increase in their use of first person plural and a drop in first person singular [5, 3]. This research has also found that certain function words are useful correlates of a variety of personality markers such as neuroticism, extraversion, openness to experience, self-esteem, and social dominance [18].

1.2 Computational stylistics

Computational stylistics [1, 10, 11, 13, 26, 28] views the full meaning of a text as much more than just the topic it describes or represents. Textual meaning, broadly construed, can include 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).

A key problem for stylistic text categorization is proper choice of textual features 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 representing the author's choice of one mode of expression from 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 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 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.

1.3 Feature choice

Language psychology and stylistic text categorization thus share a common foundation. Both extract topic-independent features from texts in order to isolate those features that best discriminate between text categories of interest. Researchers in each area have found that it is often simple features, rather than complex features, that offer the best discrimination for many categorization problems. But each area of research carries with it certain emphases from which the other might benefit. Work in language psychology focuses on identifying psychologically meaningful features which independently discriminate well between categories but does not necessarily integrate the multiplicity of discriminators into a single model which might be used to predict to which category a given text belongs. Work in text categorization focuses on the construction of predictive models but does not necessarily strive to fully understand the psychological significance of the underlying discriminators. The promise of synthesizing the two approaches is the creation of psychologically meaningful predictive models.

Most computational stylistics work to date has been based on hand-selected sets of content-independent features such as function words [20, 13, 28], parts-of-speech and syntactic structures [26], and clause/sentence complexity measures [33, 4]; also see the survey in [10]. While 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 linguistic motivation behind input feature sets that would relate features directly to stylistic concerns. 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 some feature selection). Some interesting and effective feature sets have been found in this way, such as [10, 11]; function words have also proven to be surprisingly effective on their own [17, 1, 2]. 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, of course, also aid 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 stylistic meaning. We apply here a methodology for constructing a lexicon using attribute-value taxonomies [30] based on principles of Systemic Functional Grammar (SFG) [9], 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 stylistic problems.

2 Lexical Stylistic Features

2.1 Function Words

Function words, defined as those frequent words that have primarily grammatical function in the language (such as *and*, *for*, and *the*), have proven quite useful for stylistic text classification in a variety of contexts (e.g., [17, 1, 2]). The intuition 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 a feature set is somewhat unsatisfying, as they can rarely give good insight into un-

derlying stylistic issues, thus our effort at developing more linguistically-value features for stylistic text classification.

2.2 Systemic Functional Grammar

Our linguistically-meaningful features are based on the theory of Systemic Functional Grammar (SFG), a functional approach to linguistic analysis [9]. SFG models languages as a system of choices of meanings to represent in language [15], 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, most often for text generation [14, 27] rather than analysis, due to the difficulty of complete 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[1]. More formally, each system network in this conception is a

---

[1] See [15] for a discussion of the full SFG grammar representation (allowing disjunction in entry conditions) which we simplify for computational ease.

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 conjunction 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 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 describes the main system networks which we use here for computational analysis of textual style. They are 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 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.

2.3 Cohesion

Cohesion refers to linguistic resources that enable language to connect to its larger context, both textual and extratextual [8]. 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 of automated extraction. Automated 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 [15, p. 519–528]. The three top-level options of CONJUNCTION are Elaboration, Extension, and Enhancement, defined as:

- Elaboration: Deepening the content in its context by exemplification or refocusing.

- Extension: Adding new related information, perhaps contrasting with the current information.
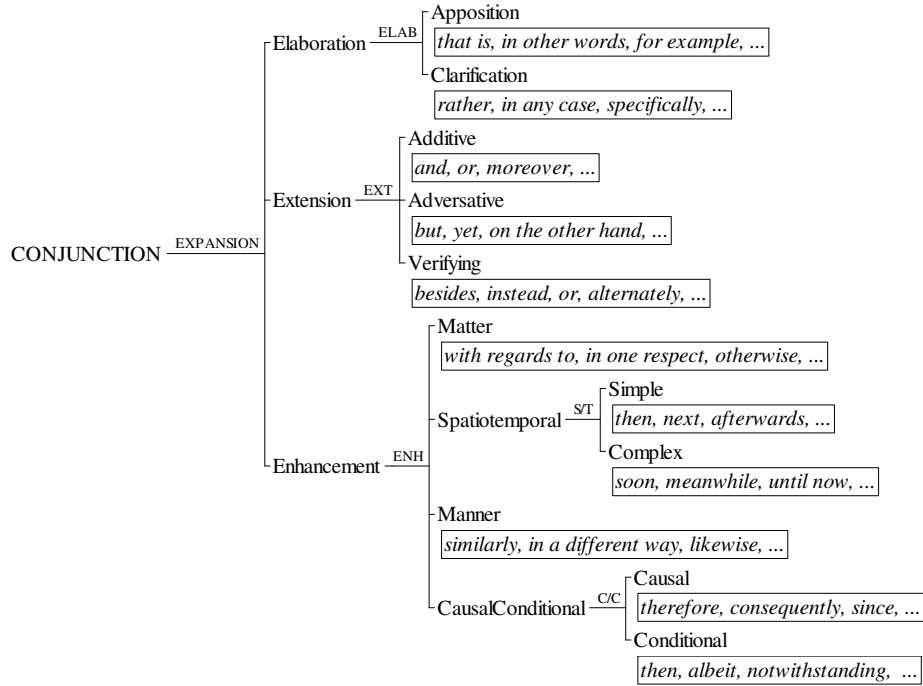
Figure 1: The CONJUNCTION system [15]. Options here are disjunctive; examples of lexical realizations for the leaves are given in italics.

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

Each of these options leads also to other options (subtypes); more detail is shown in Figure 1, and see [15, 30].

2.4 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. An important systems in SFG that address assessment is MODALITY, enabling expression of typicality and necessity of some fact or event.

The system of MODALITY enables one 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:

- 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?
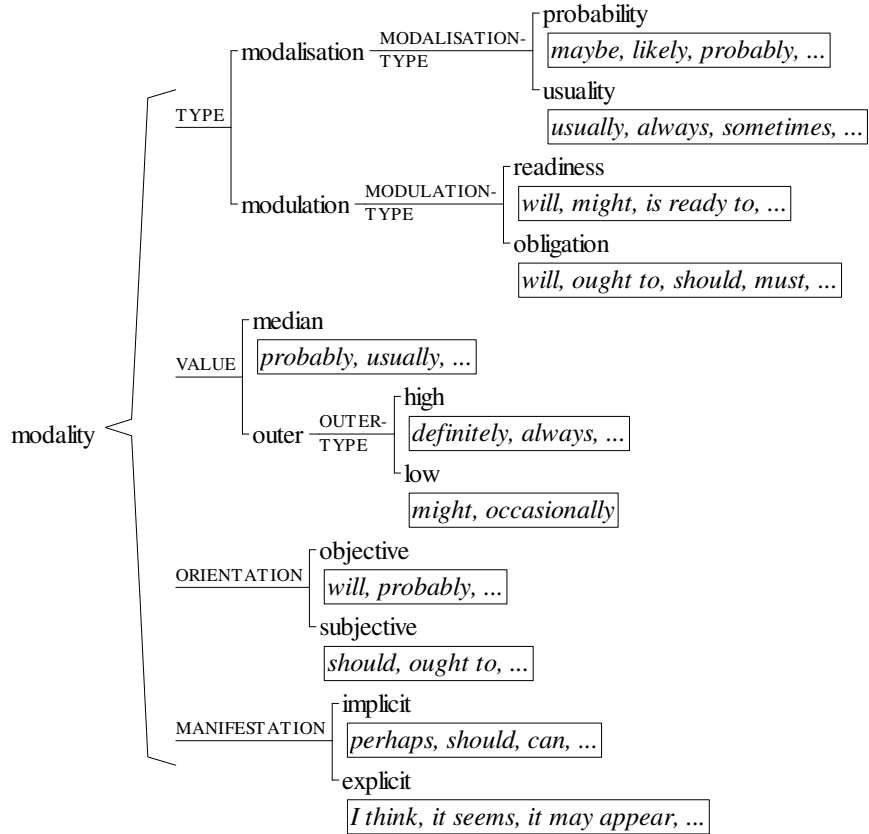
6

Figure 2: The MODALITY system networks [15], notation as above.

- 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 detailed taxonomies used for these attributes are depicted in Figure 2.

2.5 Appraisal

Finally, *appraisal* denotes how language is used to adopt or express an attitude of some kind towards some target [12]. For example, in "I found the movie quite monotonous",

7

composition-complexity
detailed, elaborate, convoluted

composition-balance
consistent, symmetrical, discordant

appreciation — APPRECIATION-TYPE

reaction-impact
amazing, compelling, dull, monotonous

reaction-quality
beautiful, elegant, awful, hideous

valuation
innovative, profound, derivative, inferior

attitude — ATTITUDE-TYPE

affect
happy, joyful, gloomy, miserable

tenacity
brave, faithful, disloyal, foolhardy

social-esteem — SOCIAL-ESTEEM-TYPE

normality
famous, lucky, popular, eccentric, obscure

capacity
clever, competent, foolish, immature

judgement — JUDGEMENT-TYPE

propriety
generous, virtuous, callous, corrupt

social-sanction — SOCIAL-SANCTION-TYPE

veracity
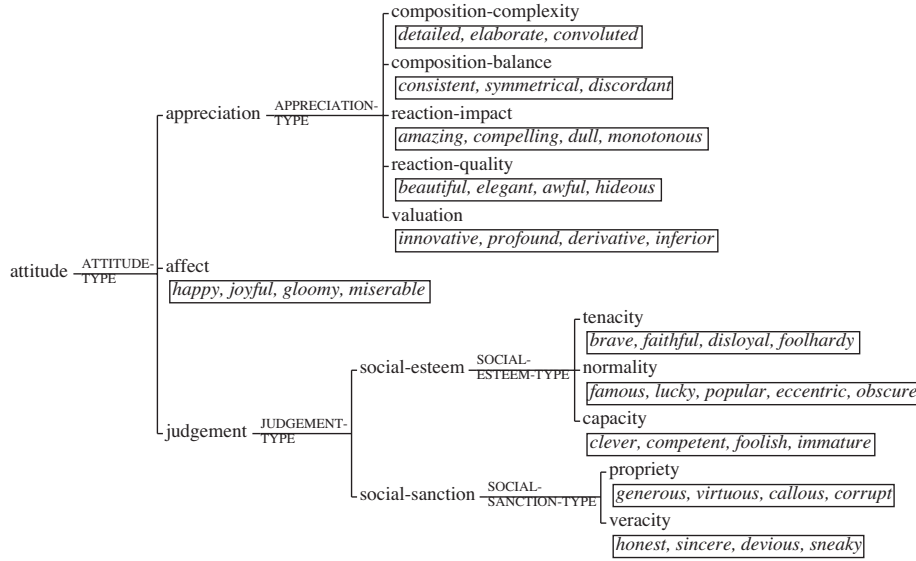*honest, sincere, devious, sneaky*

Figure 3: The Attitude network, with examples of appraisal adjectives from our lexicon.

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). 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 is shown in Figure 3.

3 Methodology

3.1 The corpus

The corpus used for this experiment was derived from essays written by students at the University of Texas at Austin between 1997 and 2003. As part of their course responsibilities, subjects (undergraduate students) wrote (inter alia) a stream-of-consciousness essay and an essay of deep self-analysis; in toto these data sets comprised 1157 and 1106 documents, respectively. Subjects were also given the NEO-FFI Five-Factor Personality Inventory [16]. Scores from the Neuroticism and Extraversion factors were used to define two binary classification task: Subjects with scores in the top third of each dimension were classed as High in that dimension, and those with scores in the bottom third classed as Low. The tasks were to use textual features to determine whether each author had High or Low neuroticism or extraversion, respectively.

3.2 Features

We empirically evaluated the use of functional lexical features for stylistic classification by applying them as well as standard function words to stylistic text classification. 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:

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 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), 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.

App: This feature set comprises, for each node n in the Attitude system:

- The relative frequency (RF_d) of node n with respect to its parent, and
- Both of $RF_d(\text{Positive}|n)$ 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).

3.3 Machine learning

In each experiment Weka's [32] implementation of the SMO learning algorithm [24] with a linear kernel was used for learning classification models. Throughout, 10-fold cross-validation was used throughout to estimate out-of-training classification accuracy.

3.4 Feature analysis: Oppositions

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 *sibling oppositions* (or simply 'oppositions') they give, where a sibling 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., $RF_d(\text{Extension}|\text{CONJUNCTION})$) is indicative of class A and CONJUNCTION/Enhancement of class B, we would have the opposition:

9

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.
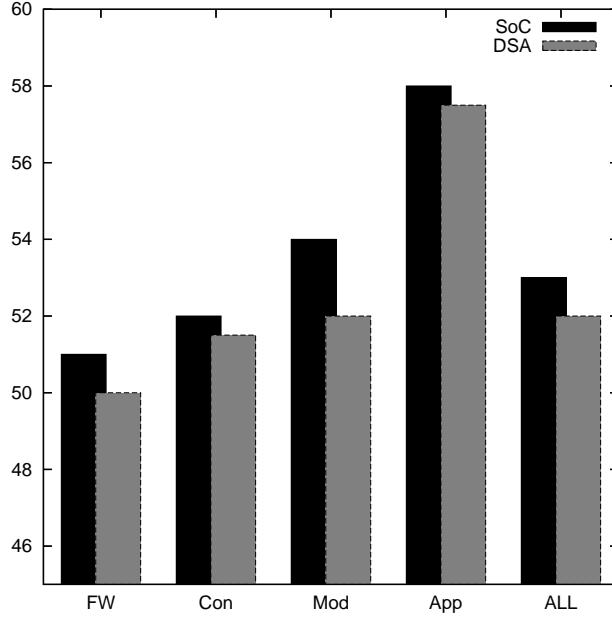
Figure 4: 10-fold cross-validation accuracy for discriminating High from Low neurotics in the stream-of-consciousness (SoC) and deep self-analysis (DSA) writing tasks.

4 Results

4.1 Neuroticism

Accuracy results for neuroticism are shown in Figure 4; the task is clearly quite difficult as the effect of personality is weak (as previously noted [23, 21]). 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). Note that this accuracy rate for determining personality type from a single short text is quite significant, as personality is difficult to diagnose without either focused questions (as in the NEO-FFI Inventory) or extended interaction (say, multiple texts written over a period of time). In any case, we expect that increasing the coverage of the Appraisal feature set to include also verbs and nouns will probably 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[2] are given in Tables 1 and 2.

First we consider the two oppositions that appear for both writing tasks[3]. 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

---

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

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

Table 1: Oppositions from the fifteen highest-ranked features indicating High and Low neuroticism in the stream-of-consciousness (SoC) 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

Table 2: Oppositions from the fifteen highest-ranked features indicating High and Low neuroticism in the deep self-analysis (DSA) 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

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 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.

4.2 Extraversion

Results for extraversion (Figure 5) are somewhat less illuminating than for neuroticism. None of the functional feature sets do as well as function words, and they even reduce accuracy overall to chance levels when added to function words. We interpret this to mean that extraversion is expressed in aspects of meaning different from Conjunction, Modality, or Appraisal.

Some tentative insight may be gleaned from examination of the most indicative function words, however, shown in Table 3. Extraverts appear, on the whole, to prefer words that suggest some relationship to norms and perhaps a sense of certainty (*immediate, am, so, being, second, normally, get, enough, very, particular*), while introverts tend to prefer
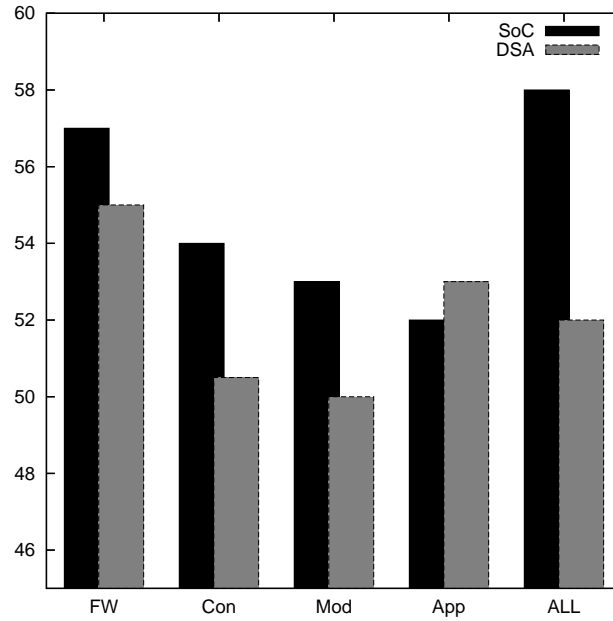
Figure 5: 10-fold cross-validation accuracy for discriminating High from Low extraverts in the stream-of-consciousness (SoC) and deep self-analysis (DSA) writing tasks.

words that suggest incompleteness or uncertainty (*perhaps, nobody, uses, try, except, getting, during, hardly*). The factor is clearly not simply certainty versus uncertainty, both because the division here is not perfect (e.g., *may* for extraverts, and *particular* for introverts), and because we would have expected Modality to do better.

5   Conclusions

We have described a method for classifying texts by personality type of the author, using functional lexical features and machine learning. For Neuroticism, our results show clearly the usefulness of such functional lexical features, in particular the Appraisal lexical taxonomy. In the case of Extraversion, results are less clear, but examination of indicative function words points a way to developing more effective features, by focusing on expressions related to norms, (in)completeness, and (un)certainty. In sum, our results confirm the utility of functional lexical features for psychological profiling (for Neuroticism), while pointing towards the need for further refinement in the feature sets and possibly the learning algorithms used (to improve overall accuracy, and interpretability for Extraversion).

Current and future work includes developing more and more extensive taxonomies for functional lexical features, as well as developing shallow parsing techniques for extracting phrases and using them in classification (preliminary work was reported in [31]). In addition, we are working on gathering similar corpora in other languages (initially Hebrew) and developing parallel feature sets, which will enable analysis of cross-lingual and cross-cultural effects.

13

Table 3: Top sixteen function words indicating High and Low extraversion in the stream-of-consciousness (SoC) and the deep self-analysis (DSA) writing assignments, respectively.

Stream-of-consciousness (SoC)		Deep self-analysis (DSA)	
Low	**High**	**Low**	**High**
perhaps	may	seven	second
outside	immediate	try	comes
nobody	anyways	self	inner
fifth	yourself	except	normally
particular	mean	getting	get
uses	am	during	enough
second	so	hardly	very
their	being	sensible	particular

Acknowledgments

References

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

[2] Shlomo Argamon and Shlomo Levitan. Measuring the usefulness of function words for authorship attribution. In *Proceedings of the 2005 ACH/ALLC Conference*, Victoria, BC, Canada, June 2005.

[3] M. Cohn, M. R. Mehl, and J. W. Pennebaker. Linguistic markers of psychological change surrounding september 11, 2001. *Psychological Science*, in press.

[4] O. de Vel. Mining e-mail authorship. In *Workshop on Text Mining, ACM International Conference on Knowledge Discovery and Data Mining*, Boston, MA, 2000.

[5] E. M. Gortner and J. W. Pennebaker. The anatomy of a disaster: Media coverage and community-wide health effects of the Texas A&M bonfire tragedy. *Journal of Social and Clinical Psychology*, in press.

[6] L. A. Gottschalk and G. Gleser. *The measurement of psychological states through the content analysis of verbal behavior*. University of California Press, Berkeley, 1969.

[7] D. T. Graham, J. A. Stern, and G. Winokur. Experimental investigation of the specificity of attitude hypothesis in psychosomatic disease. *Psychosomatic Medicine*, 20:446–457, 1958.

[8] M. A. K. Halliday and R. Hasan. *Cohesion in English*. Longman, 1976.

[9] Michael A. K. Halliday. *Introduction to Functional Grammar*. Edward Arnold, second edition, 1994.

[10] J. Karlgren. *Stylistic Experiments for Information Retrieval*. PhD thesis, SICS, 2000.

[11] Moshe Koppel, Navot Akiva, and Ido Dagan. 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, 2003.

[12] J. R. Martin and P. R. R. White. *The Language of Evaluation: Appraisal in English.* Palgrave, London, 2005. (`http://grammatics.com/appraisal/`).

[13] R. A. J. Matthews and T. V. N. Merriam. *Distinguishing literary styles using neural networks*, chapter 8. IOP publishing and Oxford University Press, 1997.

[14] C. Matthiessen and J. A. Bateman. *Text generation and systemic-functional linguistics: experiences from English and Japanese.* Frances Pinter Publishers and St. Martin's Press, London and New York, 1991.

[15] Christian Matthiessen. *Lexico-grammatical cartography: English systems.* International Language Sciences Publishers, 1995.

[16] R. R. McCrae and Jr. P. T. Costa. Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins, editor, *The five-factor model of personality: Theoretical perspectives*, pages 51–87. Guilford, New York, 1996.

[17] A. McEnery and M. Oakes. Authorship studies/textual statistics. In *Handbook of Natural Language Processing*. Marcel Dekker, 2000.

[18] M. R. Mehl and J. W. Pennebaker. The social dynamics of a cultural upheaval: Social interactions surrounding September 11, 2001. *Psychological Science*, in press.

[19] E. Mergenthaler. Emotion-abstraction patterns in verbatim protocols: A new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology*, 64:1306–1315, 1996.

[20] F. Mosteller and D. L. Wallace. *Inference and Disputed Authorship: The Federalist.* Series in behavioral science: Quantitative methods edition. Addison-Wesley, Massachusetts, 1964.

[21] J. Oberlander and A. Gill. Individual differences and implicit language: Personality, parts-of-speech and pervasiveness. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, pages 1035–1040, Chicago, 2004, 2004.

[22] J. W. Pennebaker and T. C. Lay. Language use and personality during crises: Analyses of mayor rudolph giuliani's press conferences. *Journal of Research in Personality*, 36:271–282, 2002.

[23] J. W. Pennebaker, M. R. Mehl, and K. Niederhoffer. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54:547–577, 2003.

[24] J. Platt. Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods – Support Vector Learning*. MIT Press, 1998.

[25] S. D. Rosenberg and G. J. Tucker. Verbal behavior and schizophrenia: The semantic dimension. *Archives of General Psychiatry*, 36:1331–1337, 1979.

[26] Efstathios Stamatatos, Nikos Fakotakis, and George K. Kokkinakis. Automatic text categorization in terms of genre, author. *Computational Linguistics*, 26(4):471–495, 2000.

[27] Elke Teich. *A Proposal for Dependency in Systemic Functional Grammar – Metasemiosis in Computational Systemic Functional Linguistics*. PhD thesis, University of the Saarland and GMD/IPSI, Darmstadt, 1995.

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

[29] W. Weintraub. *Verbal behavior: Adaptation and psychopathology*. Springer, New York, 1981.

[30] C. Whitelaw and S. Argamon. Systemic functional features in stylistic text classification. In *Proc. AAAI Fall Symposim on Style and Meaning in Language, Art, Music, and Design*, Washington, DC, October 2004.

[31] C. Whitelaw, N. Garg, and S. Argamon. Using appraisal taxonomies for sentiment analysis. In *Proc. Second Midwest Computational Linguistic Colloquium (MCLC 2005)*, Columbus, Ohio, May 2005.

[32] Ian H. Witten and Eibe Frank. *Data Mining: Practical machine learning tools with Java implementations*. Morgan Kaufmann, San Francisco, 2000.

[33] G. U. Yule. *Statistical Study of Literary Vocabulary*. Cambridge University Press, 1944.