A Lexical, Syntactic, and Semantic Perspective for Understanding Style in Text

Gaurav Verma, Balaji Vasan Srinivasan
BigData Experience Lab
Adobe Research, India
{gaverma, balsrini}@adobe.com

Abstract

With a growing interest in modeling inherent subjectivity in natural language, we present a linguistically-motivated process to understand and analyze the writing style of individuals from three perspectives: lexical, syntactic, and semantic. We discuss the stylistically expressive elements within each of these levels and use existing methods to quantify the linguistic intuitions related to some of these elements. We show that such a multi-level analysis is useful for developing a well-knit understanding of style – which is independent of the natural language task at hand, and also demonstrate its value in solving three downstream tasks: authors' style analysis, authorship attribution, and emotion prediction. We conduct experiments on a variety of datasets, comprising texts from social networking sites, user reviews, legal documents, literary books, and newswire. The results on the aforementioned tasks and datasets illustrate that such a multi-level understanding of style, which has been largely ignored in recent works, models style-related subjectivity in text and can be leveraged to improve performance on multiple downstream tasks both qualitatively and quantitatively.

Introduction

Modeling inherent subjectivity in natural language is of key importance for making advances in computational social science. While the notions of subjectivity pertaining to sentiment and opinion have attracted attention from computational linguists, a similar linguistic analysis of writing style has been missing from the current literature. Nonetheless, there has been a growing interest in *solving tasks* related to style in text (Peng et al. 2018; Niu and Bansal 2018; Fu et al. 2017). These approaches, however, have been limited due to their assumptions about style and its composition. They use stylistic intuitions that are linked to *differences* in style – be it genre classification (Kessler, Numberg, and Schütze 1997), author profiling (Garera and Yarowsky 2009), social relationship classification (Peterson, Hohensee, and Xia 2011), readability classification (Collins-Thompson and Callan 2005), stylized text generation (Hovy 1990; Inkpen and Hirst 2006) or style transfer (Li et al. 2018; Prabhumpe et al. 2018). These assumptions are often task-specific and do not cover all aspects of style leading to a need to fill the gap between the understanding of style and solving tasks related to it. In this work, we present a linguistically-motivated process to develop a task-independent understanding of style – that is not tied to any of the above tasks, and is general enough to encompass them.

Preprint. Under Review.

Stylistic variations in language are reflections of factors like context, author-reader dynamics, and the backgrounds of the parties involved. The influence of these factors has been analyzed in detail by psycholinguists (Semino and Culpeper 2002; Enkvist 1985). Linguistic style also deals with the prescriptive grammar associated with the aesthetics of text, as analyzed by computational linguists (Lakoff 1979; Thurmair 1990). Our exploration in this paper is centered around the computational linguistics perspective.

Earlier efforts of understanding style in text (Strunk and White 1979; DiMarco and Hirst 1988; Crystal and Davy 2016) focus on laying out *stylistic elements*, defined as the components of language that are stylistically expressive. Using this definition, we identify and discuss the elements at ***lexical*** (relating to the vocabulary of a language), ***syntactic*** (relating to the arrangement of words and phrases to create well-formed sentences) and ***semantic*** (relating to meaning in language) levels. Our qualitative analysis highlights key aspects of style that have been ignored in recent works (Peng et al. 2018; Prabhumoye et al. 2018; Jhamtani et al. 2017). We follow our qualitative analysis with a discussion of existing methods that can be used to quantify these aspects and facilitate their computational modeling. We demonstrate the value of a well-knit understanding of style by solving three downstream tasks: ***analysis of writing style*** of 5 popular English authors, ***authorship attribution*** and ***emotion prediction***. We conduct experiments on datasets that cover a diverse range of topics and domain, comprising of social media posts (Facebook and Twiiter), user reviews (IMDb), legal documents, literary books, and newswire articles (Reuters).

The rest of the paper is structured as follows: we start by discussing the proposed multi-level structure to understand style in text with several qualitative examples. In the subsequent section, we quantify representative stylistic elements to demonstrate the value of such an understanding in analyzing authors' style. In the following sections, we further use these quantifications to solve the tasks of authorship attribution and emotion prediction. We discuss various insights from our results as well as the related prior work towards the end of the paper. In the final section, we conclude the work while laying out the scope for future work.

Elements of Style in Text

Stylistically expressive elements in text can be identified at word-level (lexical), in the way sentences are structured (syntactic), and by analyzing the attributes of core-meaning that is conveyed (semantic). However, it must be noted that a style element can belong to one or more of the above categories (DiMarco and Hirst 1988). Here, we briefly describe each of the style elements and also provide examples to demonstrate the non-trivial entanglement of style and meaning in text.

Lexical Elements of style are expressed at word-level, and the stylistic variation can arise due to addition, deletion, or substitution of words. These variations can give rise to text that is characteristically different in terms of sentiment, formality, excitement etc. For example, words like *residence* and *occupied* are objective in nature, and emotionally-distant from their subjective counterparts, *home* and *busy* (Brooke and Hirst 2013a).

We also observe change in meaning and sentiment with some word-level variations: *Great food but horrible staff* vs. *Great food and awesome staff* (Li et al. 2018).

Brooke, Wang, and Hirst (2010, 2013b) enumerate such stylistic dimensions represented in lexicon as: colloquial vs. literary; concrete vs. abstract; subjective vs. objective; and formal vs. informal. For instance, while the words *residence* and *occupied* are objective, *home* and *busy* are subjective; while the word *tasty* is colloquial, *palatable* is literary.

Syntactic Elements of style are prominent in language – some examples being, *piled-up adjectives*, *detached adjectival clause*, *adjectival phrase*[1] (DiMarco and Hirst 1988).

The use of active voice is more direct and energetic than passive (Strunk and White 1979). Rhetorical theories have contrasted *loose* and *periodic* sentences – placing the most important clause at the beginning vs. placing it at the end of a sentence[2]. Such stylistic variations are well captured in the parse trees generated using probabilistic context-free grammar (PCFG) (Booth and Thompson 1973). For instance, the parse trees of loose sentences are deeper and unbalanced, while those for periodic sentences are relatively more balanced and wider (Feng, Banerjee, and Choi 2012a).

It is noteworthy that some of these syntactic style elements express themselves over multiple sentences, and are not constrained within a single sentence. For example, the use of several loose sentences in succession leads to triteness due to mechanical symmetry and a singsong effect (Strunk and White 1979). The classification of sentences as simple, compound, complex, and complex-compound and computing their statistics has facilitated identifying and differentiating between various author's styles (Feng, Banerjee, and Choi 2012b).

Semantic Elements of style can be identified by analyzing the attributes of underlying meaning that is being conveyed in a piece of text. For example, consider the two sentences: *He was not very often on time* vs. *He usually came late*. While both the sentences have a similar core-meaning, the former seems rather hesitating and noncommital, while the latter stands strong and resolute – being able to express a negative in a positive form.

Following two sentences put to contrast the vagueness and concreteness of meaning that is being conveyed: *He showed satisfaction as he took possession of his well-earned reward* vs. *He grinned as he pocketed the award* (Strunk and White 1979). It should be noted that semantic elements of style are identified by analyzing the larger meaning of the text (phrase, sentence, or paragraph), unlike lexical or syntactic elements which consider the meaning of the comprising words or the syntax of the underlying sentence.

Various stylistically expressive elements can occur in conjunction. DiMarco and Hirst present an interesting example where a detached adjectival clause, which is a syntactic element of style, can present variations that are semantically concordant and discordant: *The university attended by the President, one of the finest law schools in the country, is the alma mater of many politicians* vs. *The university attended by the President, a set of building with an architectural charm of a prison, is the alma mater of many politicians*.

Apart from these 3 elements of style, there are **surface-level style elements** that relate to length of sentences, use of punctuation, length of words, length of paragraphs, etc. For instance, use of Oxford commas is mandated by the American Psychological Association but not recommended by the Associated Press (Goldstein 1998).

Separability of Meaning and Style: Let us consider the following set of sentences:
$S1$: Chocolates can kill you.
$S2$: Chocolates, although tasty, can kill you.
$S3$: Chocolates, although palatable, can kill you.
$S4$: A period of unfavorable weather set in.
$S5$: It rained every day for a week.
From our discussion above, $S2$ and $S3$ present variation at lexical-level – $S3$ is literary while $S2$ is more colloquial. $S1$ and $S2$ express variation at syntactic-level – $S2$ has an additional detached adjective. $S4$ and $S5$ illustrate variation at semantic-level – one can notice that the core-meaning in $S4$ is vague while $S5$ is definite and specific.

These examples illustrate that style is not merely a means of expressing meaning, but plays an active role in bringing about meaning. While there are variations in style that are separable from meaning (for example, the syntactic switch in style between active and passive voice: *I shall always remember my first visit to Boston* vs. *My first visit to Boston will always be remembered*), the disentanglement of style

[1] **Piled-up adjectives**: The *gigantic green* dragon felt bad. **Detached adjectival clause**: He, *being a recluse*, often quietly excused himself. **Adjectival phrase:** The movie was *not too terrible*.

[2] **Loose sentence**: "*These algorithms can not deal with words for which classifiers have not been trained*"; **Periodic sentence**: "*For processing free texts hand-crafted grammars are neither practical nor reliable.*" (Feng, Banerjee, and Choi 2012a)

Level	Stylistic Element	Abraham Lincoln	Mark Twain	Oscar Wilde	Rudyard Kipling	Charles Dickens
Surface	Avg # of words in a sentence ($\mu \pm \sigma$)	**30.88** $\pm$ 4.16	**23.22** $\pm$ 2.64	20.35 $\pm$ 2.91	**19.29** $\pm$ 3.12	**26.77** $\pm$ 3.91
	Avg # of commas; semicolons; colons	**1.28**; 0.11; 0.03	1.14; 0.31; 0.08	0.82; 0.05; 0.01	0.93; 0.57; 0.02	**1.42**; 0.19; 0.03
	Avg # of sentences in a para. ($\mu \pm \sigma$)	**3.52** $\pm$ 0.63	5.59 $\pm$ 0.42	5.19 $\pm$ 0.39	5.35 $\pm$ 0.47	4.53 $\pm$ 0.57
Lexical	Literary vs. colloquial words	**0.91** vs. 0.09	0.73 vs. 0.27	0.69 vs. 0.31	0.61 vs. 0.39	0.75 vs. 0.25
	Abstract vs. concrete words	**0.71** vs. 0.29	0.64 vs. 0.36	0.58 vs. 0.42	**0.39** vs. **0.61**	0.54 vs. 0.46
	Subjective vs. objective words	0.57 vs. 0.43	0.58 vs. 0.42	0.68 vs. 0.32	**0.48** vs. **0.52**	0.57 vs. 0.43
	Formal vs. informal words	**0.87** vs. 0.13	0.74 vs. 0.26	0.70 vs. 0.30	0.57 vs. 0.43	0.62 vs. 0.38
Syntactic	Fraction of simple sent.	**0.09**	**0.18**	0.17	**0.26**	**0.13**
	Fraction of compound sent.	0.21	0.23	0.24	0.24	0.21
	Fraction of complex sent.	**0.31**	0.27	0.26	0.25	**0.30**
	Fraction of complex-compound sent.	**0.35**	0.28	0.26	0.23	**0.29**
	Fraction of loose sentences	**0.17**	0.09	0.07	**0.03**	0.16
	Fraction of periodic sentences	**0.12**	0.14	0.11	**0.08**	0.12

Table 1: *Quantification of stylistic elements at different levels, for author analysis.* The presented values are averages computed over all 10 documents for each author. For surface-level analysis, we report the averages and standard deviations ($\mu \pm \sigma$). For lexical analysis, the values for a vs. b are equal to $\frac{\text{\# of words in } a}{\text{\# of words in } a \text{ or } b}$ and $1 - \frac{\text{\# of words in } a}{\text{\# of words in } a \text{ or } b}$ respectively. For syntactic analysis, the sum of all fractions corresponding to simple, compound, complex, and complex-compound is ≤ 1, owing to sentences that do not fall into either of those categories – *What an idiot!* being one such example of an incomplete sentence. The values in **bold** are of interest and have been mentioned in the text.

and meaning can only be applied to a few stylistic elements and the process becomes increasingly complex as we progress from lexical, to syntactic, to semantic variations.

Recent approaches in style-related tasks assume independence between meaning and style (Li et al. 2018; Jhamtani et al. 2017; Prabhumoye et al. 2018) restricting the understanding of style to a means of reflecting already existing meaning (Tikhonov and Yamshchikov 2018; Dai et al. 2019). For example, recent style transfer approaches involve generating realization $\mathbf{x}_2$ tuned to style $\mathbf{y}_2$, from realization $\mathbf{x}_1$ possessing style $\mathbf{y}_1$, by learning an auto-encoder model, that infers $\mathbf{x}_1$'s latent meaning $\mathbf{z} \sim p(\mathbf{z}|\mathbf{x}_1, \mathbf{y}_1)$ and then generates the transferred counterpart from $p(\mathbf{x}_2|\mathbf{z}, \mathbf{y}_2)$ (Shen et al. 2017). This approach is useful where style $\mathbf{y}_1$ can be separated from $\mathbf{x}_1$ to represent the meaning $\mathbf{z}$, *and* the *same* meaning can be used to generate realization $\mathbf{x}_2$ tuned to $\mathbf{y_2}$. While their approach can handle most of the lexical-level variations (see $S2$ vs. $S3$), they cannot handle other variations at syntactic (see $S1$ vs. $S2$) or semantic-level ($S4$ vs. $S5$). Recently, registering these shortcomings of existing work, Dai et al. (2019) aim to perform the task of style transfer without learning disentangled latent representations. Our work is a step towards understanding linguistic aspects of style and can be used to expand on the approaches to solve style-related natural language tasks.

Authors' Style Analysis

We quantify a few representative stylistic elements[3] at each of these levels and analyze the writings of 5 popular English authors (Abraham Lincoln, Mark Twain, Oscar Wilde, Rudyard Kipling and Charles Dickens) from the Gutenberg dataset (Lahiri 2014). The subset of Gutenberg dataset that we consider for analyzing the writing style of authors comprises of 50 published books (10 books per author). Each of the books contain, on average, $\sim 1,000$ sentences and the entire corpus of 50 books contains around $25,000$ unique words (such that their frequency of occurrence is greater than or equal to 3 in the entire corpus of 50 books). In the supplementary document, for the ease of reproducibility, we provide the author-wise list of books and *representative* writing samples from the five authors under consideration.

At surface-level, we report the average number of words in a sentence, average number of commas, semicolons and colons in a sentence, and average number of sentences in a paragraph. For quantifying the lexical elements, we use a list of seed words for each of the following eight categories: subjective, objective, concrete, abstract, literary, colloquial, formal and informal (Brooke and Hirst 2013b). Following Brooke and Hirst (2013b), we compute normalized pointwise mutual information index (PMI) to obtain a raw style score for each dimension, by leveraging co-occurrences of words in the entire corpus. The raw scores are normalized to obtain style vectors for every word, followed by a transformation of style vectors into k-Nearest Neighbor (kNN) graphs, where label propagation is applied. It is worth noting that the eight original dimensions lie on the two extremes of four different spectrums, i.e., subjective-objective, concrete-abstract, literary-colloquial, and formal-informal. We compute averages across a given author's corpus. The averages, in the range $[0, 1]$, denote the author's tendency to use subjective, concrete, literary, or formal words, in contrast to using objective, abstract, colloquial, or informal words, respectively, as evidenced in their literary works.

For syntactic analysis, we use sentence structure identification algorithms proposed by Feng, Banerjee, and Choi (2012a), and compute the fractions of simple, compound, complex, and complex-compound sentences. The numbers reported in Table 1 are fractions of all sentences in the concerned author's corpus. We also quantify the fraction of sentences that are identified as periodic and loose.

For semantic analysis, in absence of clear approaches for quantification, we observe the output of a knowledge parser, *K-Parser* (Sharma et al. 2015), and report most frequent entities and their semantic roles (Palmer, Gildea, and Kingsbury 2005).

The statistics in Table 1 indicate that the quantification aligns with several qualitative observations pertaining to the

[3]We do not aim to quantify *all* the discussed stylistic elements. The primary goal is to provide empirical evidence to substantiate the claim that such a multi-level understanding of style is of value in solving downstream tasks. Quantification of some of the discussed stylistic elements, especially the ones at semantic-level, is an open problem in itself and calls for further research.

authors. For instance, writings of Abraham Lincoln involve statements of political significance and hence, are carefully structured (see the supplementary material for representative writing samples). At surface-level, this is reflected in longer sentences, extensive use of commas, and fewer sentences in a paragraph. At lexical-level, the use of abstract words like *freedom*, *respect*, *passion* stands out. It is also notable that since loose and periodic sentences typically have a non-simple syntactic structure, Lincoln has a significantly larger fraction of sentences that are either identified as loose or periodic sentences. The use of words that are more literary than colloquial and more formal than informal, is also quite prominent. At syntactic-level, Lincoln's sentences are structured complicatedly and very few sentences have a simple syntactic structure. Lincoln often referred to a nation as a living entity, which is in turn observed as *nation* being one of the frequently used entities with its semantic role being `:alive entity`. Along similar lines, the semantic role for the entity *law* was frequently found to be `:impelling agent`.

Mark Twain, while having similar lexical-level polarities as Lincoln, uses relatively simpler sentences. Oscar Wilde, known for his satire on contemporary culture (Ellmann 1988), frequently uses the entity *people* in his writings, where the two frequently associated semantic roles are `:object of affection` and `:object of disaffection`. It is interesting to note that the characteristically long sentences that are attributed to Charles Dickens (Hobsbaum 1998) are also captured in this multi-level analysis of style – significantly higher number of words in a sentence, with prominent use of punctuation for conjunctions (i.e., commas), and a higher tendency to use complex and complex-compound sentences.

The stylistic *variations* in the writings are well captured across all the levels. For example, Rudyard Kipling, well-known for short stories and classics of children's literature (Wilson 1979), has a higher tendency of forming short sentences with simple syntactic structure than other authors – which is in turn also reflected in comparatively lower fractions of loose/periodic sentences. Additionally, Kipling is the only author to use more concrete words like *gongs*, *rockets*, *torch* etc. and less abstract words like *suffer*, *freedom*, etc.

These observations reinforce that writing style is a compound factor of several stylistic elements and can be identified at multiple levels in text, validating the need of a multi-level analysis. The *differences* in writing style of authors that are observed across these levels further strengthen the support for such a multi-level analysis.

Authorship Attribution

To further illustrate the value of a task-independent understanding of style, we use the quantified style elements to solve the authorship attribution task – the task of identifying the author of a document (Love 2002). We use the methods by Sari, Stevenson, and Vlachos (2017; 2018) as the baseline, and analyze the effects of adding our multi-level stylistic features. We use four datasets – Judgement (Serioussi, Smyth, and Zukerman 2011), CCAT10, CCAT50 (Stamatatos 2008), and IMDb62 (Seroussi, Zukerman, and Bohnert 2010) which cover a range of characteristics in terms of number of authors, topic, and document length (Sari, Stevenson, and Vlachos 2018) (see Table 2 for details). We concatenate our quantified stylistic features at lexical, surface and syntactic level (4, 3, and 6 in number, respectively; see Table 1) with the features designed by Sari, Stevenson, and Vlachos (2018). The original features that Sari, Stevenson, and Vlachos use are designed to capture authors' writing style and topical preference.

Characteristics	**Judgement**	**CCAT10**	**CCAT50**	**IMDb62**
genre	legal	newswire		movie reviews
# authors	3	10	50	62
# of documents	1,342	1000	5000	79,550
avg chars / doc	11,957	3,089	3,058	1,401
avg words / doc	2,367	580	584	288

Table 2: Data statistics for the authorship attribution task

Baselines: Sari, Vlachos, and Stevenson (2017) propose to represent a document as a bag of character-based n-gram features and learn the continuous representation of each feature jointly with the classifier in a shallow feed-forward neural network. Following this work, in 2018 they extend their character-based model by incorporating a combination of style and content related features as auxiliary features represented in discrete form. Their style-based features include features like average word length, number of short words, frequency of function words, occurrence of punctuation, etc., while their content-based features include frequency of uni/bi/tri-grams of common words. These auxiliary features provide additional information related to the dataset characteristics. We concatenate our proposed multi-level stylistic features to these auxiliary features and analyze their efficacy. It is worth noting that to isolate the effects of modeling changes and input feature changes, we keep the hyperparameters same as those in the baseline models. We also compare the performance of our proposed approaches with other state-of-the-art models (Seroussi, Zukerman, and Bohnert 2014; Escalante, Solorio, and Montes-y Gómez 2011; Parikh, Venkataram, and Kalita 2018). For comparisons in Table 3 and 4, we use the same data preprocessing techniques and model hyperparameters as described in their work by Sari, Stevenson, and Vlachos (2017; 2018)[4]. Please refer to Sari, Vlachos, and Stevenson (2017; 2018) for further details.

Table 3 summarizes the effect of using proposed stylistic features along with existing features with respect to the continuous character n-grams model (Sari, Vlachos, and Stevenson 2017). In Table 3, we report the average accuracy over 20 different experimental runs. Additionally, in Table 4 we report the change in accuracy brought by adding *(a)* new stylistic features at individual levels and *(b)* the stylistic levels across *all* levels. We also indicate the statistical significance (*t-test*) of presented results in Table 4.

It can be noted from Table 3 that the inclusion of proposed multi-level stylistic features improves the performance of

[4]URL for reproducing the baselines: *https://github.com/yunitata/continuous-n-gram-AA* and *https://github.com/yunitata/coling2018*

Models	Judgement	CCAT10	CCAT50	IMDb62
SVM with bag of local histogram (Escalante, Solorio, and Montes-y Gómez 2011)	–	**86.40%**	–	–
Token SVM (Seroussi, Zukerman, and Bohnert 2014)	91.15%	–	–	91.52%
Using topic models (Seroussi, Zukerman, and Bohnert 2014)	93.64%	–	–	91.79%
Document embeddings based on textual style (Parikh, Venkataram, and Kalita 2018)	–	63.80%	76.60%	89.90%
†Continuous character n-grams + content & style (Sari, Vlachos, and Stevenson 2017; Sari, Stevenson, and Vlachos 2018)	91.51%	76.20%	72.88%	95.93%
Continuous character n-grams + content & style + **New features**	**94.44%**	80.07%	**77.75%**	**97.89%**

Table 3: *Performance of our proposed approach on the authorship attribution task.* We concatenate multi-level stylistic features with the auxiliary features of the baseline† and compare the classification accuracies.

Features	Judgement	CCAT10	CCAT50	IMDb62
baseline features†	91.51%	76.20%	72.88%	95.93%
(+) lexical	+1.17	+2.17	+2.87	+0.93
(+) surface	+0.09*	+0.13	+0.07	+0.11
(+) syntactic	+1.57	+1.29	+1.21	+0.87*
(+) all	+2.91	+3.87	+4.87	+1.96

Table 4: *Stylistic feature ablation results for authorship attribution task.* + denotes % increase over baseline† due to addition of proposed stylistic features. Underlined values are statistically significant with $p < 0.001$, while those with * are significant with $p < 0.01$.

existing state-of-the-art models on Judgement, CCAT50, and IMDb62. More importantly, a further analysis in Table 4 shows that the addition of proposed stylistic features to the baseline features, results in improvement of classification accuracies across the four datasets. The improved performance due to stylistic features at individual levels indicates their ability to capture new notions of style and the significant increase when *all* the style elements are used together, reinforces the need of a multi-level stylistic analysis. As a sidenote, in Table 4, it can observed that addition of surface-level stylistic elements does not improve the classification accuracy as much as addition of lexical and syntactic elements do. This can be attributed to the fact that most of the existing style-based features in the baseline can be identified as surface-level features, whereas very few can be identified as lexical or syntactic.

Emotion Prediction

We now illustrate the value of a multi-level analysis of style in solving the task of *fine-grained* emotion classification. While emotion can be classified on a discrete level (e.g., happy, sad, excited, etc.) we focus on a fine-grained classification using valence, arousal, and dominance values (Strapparava and Mihalcea 2007; Buechel and Hahn 2017). The manner in which meaning is conveyed influences the emotion it evokes in readers of a given text (Wise et al. 2009; Kao and Jurafsky 2012). While the relationship between content and emotion has been studied extensively (Subasic and Huettner 2001; Neviarouskaya, Prendinger, and Ishizuka 2011; Kantrowitz 2003), owing to availability of language-specific resources (Mohammad, Kiritchenko, and Zhu 2013; Mohammad 2018), little research has been done to study the relationship between style and emotion (Kao and Jurafsky 2012). The motivation for considering the task of emotion prediction in this work is twofold: *(a)* analyze the role of stylistic aspects of text in predicting emotion, and *(b)* analyze the value of having a multi-level stylistic representation as proposed in this work.

We consider the method proposed by Akhtar et al. (2018) as a baseline and concatenate our proposed multi-level stylistic features with their existing features. The baseline and the proposed modification is evaluated on two standard datasets for emotion classfication: the EmoBank dataset (Buechel and Hahn 2017) and the Facebook posts dataset (Preoţiuc-Pietro et al. 2016). The EmoBank dataset comprises of $10,062$ tweets across multiple domains (e.g. blogs, news headlines, fiction etc.). Each tweet has three scores representing valence, arousal and dominance of emotion on a continuous range of 1 to 5. The Facebook posts dataset contains $2,895$ social media posts that are annotated on a nine-point score with valence and arousal scores by two psychologically trained annotators. To ensure consistency while comparing results with the baselines, for experiments, we adopt 70-10-20 split for training, validation and testing, respectively. As stated in the work by Akhtar et al. (2018), we perform 10-fold cross-validation for the evaluation. We also use the same training and evaluation setup, along with same model hyperparameters, to ensure meaningful comparisons. For more implementations details, please refer to Akhtar et al. (2018).

Baselines: Akhtar et al. (2018) propose a multi-task ensemble that combines the learned representations of three independently trained deep learning models (i.e., a Convolutional Neural Network (CNN), a Long Short Term Memory (LSTM), and a Gated Recurrent Unit (GRU) network) and a hand-crafted feature vector that comprises of features like word and character tf-idf, lexicon-based sentiment scores, count of positive and negative words, etc. The multi-task ensemble is essentially a multi-layer perceptron (MLP) with two shared hidden layers and two task-specific hidden layers that cater to the specific need of individual tasks. Their motivation of solving the three regression problems (one for each valence, arousal and dominance) in a multi-task setup arises from the intuition that these related tasks can help the joint-model learn effectively from shared representations while achieving better generalization. To assess the role of our proposed stylistic features in predicting emotion, we concatenate them with the handcrafted features of Akhtar et al. For comparison with other existing state-of-the-art methods, we also include the performance of the System proposed by Preoţiuc-Pietro et al. (2016).

As it can be observed from results presented in Table 5, the addition of new multi-level stylistic features leads to significant improvement in the Pearson correlation coefficient, over multiple baselines. Pearson correlation coefficient measures the linear correlation between the actual and predicted scores and has been used extensively in prior art (Mohammad and Bravo-Marquez 2017; Preoţiuc-Pietro et al. 2016).

In Table 5, we quantify the improvement brought by incorporating stylistic elements at individual levels. The aver-

Models	EmoBank			FB Post	
	Val	Aro	Dom	Val	Aro
System (Preoţiuc-Pietro et al. 2016)	–	–	–	0.650	**0.850**
CNN (C) (Akhtar et al. 2018)	0.567	0.347	0.234	0.678	0.290
LSTM (L)	0.601	0.337	0.245	0.671	0.324
GRU (G)	0.569	0.315	0.243	0.668	0.313
Ensemble (CLG)	0.618	0.365	0.263	0.695	0.336
‡Ensemble (CLG + Old features)	0.635	0.375	0.277	0.727	0.355
(+) lexical	$\underline{+0.026}$	$\underline{+0.003}$	$+0.017^*$	$\underline{+0.010}$	$\underline{+0.043}$
(+) surface	-0.003^*	$\underline{-0.005}$	$\underline{-0.001}$	-0.006^*	$\underline{+0.008}$
(+) syntactic	$\underline{+0.012}$	$\underline{+0.006}$	$+0.002^*$	$+0.004^*$	$\underline{+0.009}$
(+) all	$\underline{+0.039}$	$\underline{+0.007}$	$+0.020^*$	$\underline{+0.009}$	$\underline{+0.062}$
Ensemble (CLG + Old & **New features**)	**0.674**	**0.382**	**0.297**	**0.736**	0.417

Table 5: *Performance on the emotion prediction task.* We concatenate multi-level stylistic features with handcrafted features and compare Pearson coefficient correlation. $+$ denotes increase over baseline‡ due to addition of proposed stylistic features. Underlined values are statistically significant with $p < 0.001$, while those with * are significant with $p < 0.01$.

age change in Pearson coefficient correlation, over 20 different experimental runs, is reported in Table 5 along with the statistical significance (*t-test*) of reported results. As it was the case with the task of author attribution (see Table 4), inclusion of *all* multi-level stylistic features allows the model to capture new notions of style and reinforces the need of a multi-level stylistic analysis. Additionally, referring back to our motivation of choosing the emotion prediction task, we provide empirical evidence that stylistic aspects do correlate with valence, arousal and dominance values. The questions around causal *significance* and *extent* (Pearl 2010; Wang and Blei 2018) of stylistic aspects towards evoked emotion are yet to be answered, and are left as a part of future work. However, we expect the interpretability of the proposed stylistic features to aid in establishing causal relationships.

Discussion of Results

To illustrate the value of the proposed multi-level representation of style, we focused on three tasks: authors' style analysis, authorship attribution and emotion prediction. Given this, it becomes essential to emphasize that this work does not aim to propose novel approaches to any of the aforementioned tasks. The primary aim of the work is an effort to establish a structured multi-level understanding of style in text that can facilitate in better modeling of style. We substantiate the value of such an understanding by giving empirical evidences in Table 1, 3, 4 and 5.

To summarize the empirical evidences, by quantifying and analyzing the writing style of 5 English authors, we demonstrate that the proposed stylistic features provide interpretable and coherent insights about an author's style. When the proposed multi-level stylistic features are added in a simplistic way to solve the tasks of authorship attribution and emotion prediction, they further improve the performance of existing state-of-the-art approaches. Specific to the task of emotion prediction, we also demonstrate that stylistic aspects of text have a correlation with the emotion it evokes in its readers.

An interesting aspect that is highlighted by solving the tasks of authorship attribution and emotion prediction is the varying extent to which stylistic elements at different levels contribute towards solving the task. For instance, in Table 5, the lexical and syntactic-level elements of style add significantly more value to the task of emotion prediction than surface-level elements. This claim is substantiated as we note that the original handcrafted features used in the baseline are devoid of any stylistic features whatsoever. The proposed structure provides more holistic interpretability while modeling style to solve related tasks.

Related Work

In this section we provide a comprehensive description of prior related work. We start by discussing in detail the work that aims to computationally model style in text. This is followed by a brief discussion of existing approaches to analyze the style of authors and solve the task of authorship attribution and emotion prediction.

Understanding Style in Text: As mentioned earlier, while there is a recent focus on solving style-related natural language tasks (Li et al. 2018; Prabhumoye et al. 2018; Jhamtani et al. 2017; Shen et al. 2017), there has been a decline in efforts that aim to identify what style constitutes and provide a holistic task-independent understanding of it (Tikhonov and Yamshchikov 2018; Crystal and Davy 2016). Given the recent advancements in machine learning and data-driven approaches in style-related problems, it is imperative that we look back to align our understanding of style to work with and aid recent methods.

Efforts to understand style range back to the work by DiMarco and Hirst (1988) where they provide a grammar of style while translating realizations from one language to the other. To do so, they introduce the notion of *internal stylistics* of source and target languages and a method to map these internal notions of style. The internal stylistics of a language relate to its linguistic characteristics and capture aspects like abstraction, dynamism, clarity, and formality. However, given the recent standing of the larger field, it is difficult to leverage these linguistic-based internal stylistics to aid data-driven computational models. Similar problems arise with other prior works (Brewer and Hay 1984; Semino and Culpeper 2002; Freeborn 1996) where the extension of linguistic intuitions to currently prevalent modeling approaches is non-trivial.

More recently, there have been efforts to quantify stylistic features to enable style-based text categorization (Koppel, Akiva, and Dagan 2003; Argamon-Engelson, Koppel, and

Avneri 1998). Since text categorization itself encompasses several downstream tasks (e.g., sentiment analysis, genre classification, authorship attribution, etc.) there is a tendency to define style – and consequently, the stylistic features – in a task-specific manner (Tikhonov and Yamshchikov 2018). It is also notable that the definition of stylistic features are inconsistent among prior works. For instance, Brooke and Hirst (2013a; 2013b) focus primarily on lexical-level stylistic aspects while Feng, Banerjee, and Choi (2012b) focus on aspects of style at syntactic-level.

Our proposed work adds to the current research by providing a structured understanding of stylistic elements at surface, lexical, syntactic and semantic-level. We also demonstrate that such an understanding, which is rooted in linguistically rich intuitions, can be used to obtain a multi-level representation of style which can be further used to improve the performance of state-of-the-art data-driven approaches across multiple tasks.

Next, we discuss the prior works that aim to solve the aforementioned tasks by taking stylistic aspects of text into account.

Authors' Style Analysis: There are several works that aim to quantify stylistic linguistic intuitions to analyze the writing style of well known authors (McCarthy et al. 2006; Peng and Hengartner 2002; Forgeard 2008). They leverage features ranging from cohesion measures[5] to reading difficulty. They often demonstrate the value of their identified features by relating qualitative insights with quantified representations. Taking motivation from here, we demonstrate the value of our proposed stylistic representation by building a coherent understanding (across multiple levels) of writing style of 5 well known English authors.

Authorship Attribution: Recent approaches to the task of authorship attribution make use of a mix of content and style-based features (Sari, Stevenson, and Vlachos 2018). While the content-based features are majorly character or word-based n-grams (Sari, Vlachos, and Stevenson 2017; Kešelj et al. 2003), the style-based features include shallow features like function word frequencies, count of *hapax legomena*, etc., and deep linguistic features like context free production frequencies and semantic relationship frequencies (Gamon 2004). Addition of our proposed stylistic representation facilitates modeling of newer notions of style in a structured and interpretable manner and improves the performance of existing state-of-the-art models.

Emotion Prediction: Features like count of positive and negative words (Wiebe and Mihalcea 2006), count of words matching each emotion from the NRC Word-Emotion Association Lexicon (Mohammad and Turney 2013), word and character tf-idf have been extensively used for predicting emotion (Akhtar et al. 2018; Preoţiuc-Pietro et al. 2016). However, the efforts that aim to analyze the influence of stylistic aspects on emotion have been scarce (Kao and Jurafsky 2012). We demonstrate that linguistically motivated stylistic features are not only correlated with emotion, but also help in improving the performance of existing emotion prediction approaches.

[5]Cohesion is the grammatical and lexical linking within a text or sentence that holds a text together and gives it meaning. For instance, the linking that enables us to understand the reply in the following conversation: *(A)* "Where are you going?" *(B)* "To dance."

Conclusion and Emerging Directions

Understanding style is an important aspect of modeling inherent subjectivity in text. We presented a linguistically-motivated process to understand and qualify stylistic aspects of text at lexical, syntactic, and semantic-level. Using existing methods to quantify these style-related linguistic intuitions, we analyzed the writing style of 5 authors, and solved the task of authorship attribution and emotion prediction. We demonstrated that the style of an author is a compound factor of various stylistic elements, validating the need of a multi-level analysis of style. We also improve the performance of existing state-of-the-art approaches to authorship attribution and emotion prediction task by modeling style in a structured and interpretable manner. To strengthen the empirical evidences, we conducted experiments on datasets that contained text from diverse topics and domains (social media posts, literary texts, legal documents and movie reviews).

Such a multi-level style analysis, by being able to incorporate broader notions of style in a more structured and interpretable manner, can aid in understanding the causal influence of style towards psycholinguistic concepts like formality, sentimentality, politeness, etc., by deconfounding other potential causes like topic. In future, we aim to analyze the influence of style on psycholinguistic concepts in further detail.

References

Akhtar, M.; Ghosal, D.; Ekbal, A.; and Bhattacharyya, P. 2018. A multi-task ensemble framework for emotion, sentiment and intensity prediction. *arXiv preprint:1808.01216*.

Argamon-Engelson, S.; Koppel, M.; and Avneri, G. 1998. Style-based text categorization: What newspaper am i reading. In *Proc. of the AAAI Workshop on Text Categorization*.

Association, A. P. 2010. *Concise rules of APA style*. American Psychological Association.

Booth, T. L., and Thompson, R. A. 1973. Applying probability measures to abstract languages. *IEEE transactions on Computers* (5):442–450.

Brewer, W. F., and Hay, A. E. 1984. Reconstructive recall of linguistic style. *Journal of Verbal Learning and Verbal Behavior* 23(2):237–249.

Brooke, J., and Hirst, G. 2013a. Hybrid models for lexical acquisition of correlated styles. In *IJCNLP*.

Brooke, J., and Hirst, G. 2013b. A multi-dimensional bayesian approach to lexical style. In *NAACL-HLT*.

Brooke, J.; Wang, T.; and Hirst, G. 2010. Automatic acquisition of lexical formality. In *23rd International Conference on Computational Linguistics*.

Buechel, S., and Hahn, U. 2017. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *EACL*.

Collins-Thompson, K., and Callan, J. 2005. Predicting reading difficulty with statistical language models. *Journal of the American Society for Information Science and Technology*.

Crystal, D., and Davy, D. 2016. *Investigating english style*.

Dai, N.; Liang, J.; Qiu, X.; and Huang, X. 2019. Style transformer: Unpaired text style transfer without disentangled latent representation. *arXiv preprint:1905.05621*.

DiMarco, C., and Hirst, G. 1988. Stylistic grammars in language translation. In *COLING*.

Ellmann, R. 1988. *Oscar wilde*. Vintage.

Enkvist, N. E. 1985. Introduction: Stylistics, text linguistics, and composition. *Text-Interdisciplinary Journal for the Study of Discourse* 5(4):251–268.

Escalante, H. J.; Solorio, T.; and Montes-y Gómez, M. 2011. Local histograms of character n-grams for authorship attribution. In *ACL*.

Feng, S.; Banerjee, R.; and Choi, Y. 2012a. Characterizing stylistic elements in syntactic structure. In *Joint Conference on EMNLP and CONLL*.

Feng, S.; Banerjee, R.; and Choi, Y. 2012b. Syntactic stylometry for deception detection. In *ACL*.

Forgeard, M. 2008. Linguistic styles of eminent writers suffering from unipolar and bipolar mood disorder. *Creativity Research Journal* 20(1):81–92.

Freeborn, D. 1996. What is style? In *Style*. Springer. 1–7.

Fu, Z.; Tan, X.; Peng, N.; Zhao, D.; and Yan, R. 2017. Style transfer in text: Exploration and evaluation. *arXiv preprint:1711.06861*.

Gamon, M. 2004. Linguistic correlates of style: authorship classification with deep linguistic analysis features. In *COLING*.

Garera, N., and Yarowsky, D. 2009. Modeling latent biographic attributes in conversational genres. In *ACL*.

Goldstein, N. 1998. *The Associated Press Stylebook and Libel Manual. Fully Updated and Revised.* ERIC.

Hobsbaum, P. 1998. *A reader's guide to Charles Dickens*. Syracuse University Press.

Hovy, E. H. 1990. Pragmatics and natural language generation. *Artificial Intelligence* 43(2):153–197.

Inkpen, D., and Hirst, G. 2006. Building and using a lexical knowledge base of near-synonym differences. *Computational linguistics* 32(2):223–262.

Jhamtani, H.; Gangal, V.; Hovy, E.; and Nyberg, E. 2017. Shakespearizing modern language using copy-enriched sequence-to-sequence models. *arXiv preprint:1707.01161*.

Kantrowitz, M. 2003. Method and apparatus for analyzing affect and emotion in text. US Patent 6,622,140.

Kao, J., and Jurafsky, D. 2012. A computational analysis of style, affect, and imagery in contemporary poetry. In *NAACL-HLT Workshop on Computational Linguistics for Literature*.

Kešelj, V.; Peng, F.; Cercone, N.; and Thomas, C. 2003. N-gram-based author profiles for authorship attribution. In *Pacific association for computational linguistics*.

Kessler, B.; Numberg, G.; and Schütze, H. 1997. Automatic detection of text genre. In *EACL*.

Koppel, M.; Akiva, N.; and Dagan, I. 2003. A corpus-independent feature set for style-based text categorization. In *Workshop on Computational Approaches to Style Analysis and Synthesis*.

Lahiri, S. 2014. Complexity of Word Collocation Networks: A Preliminary Structural Analysis. In *Student Research Workshop at the 14th EACL*.

Lakoff, R. T. 1979. Stylistic strategies within a grammar of style. *Annals of the New York Academy of Sciences*.

Li, J.; Jia, R.; He, H.; and Liang, P. 2018. Delete, retrieve, generate: A simple approach to sentiment and style transfer. *arXiv preprint:1804.06437*.

Love, H. 2002. *Attributing authorship: An introduction*.

McCarthy, P. M.; Lewis, G. A.; Dufty, D. F.; and McNamara, D. S. 2006. Analyzing writing styles with coh-metrix. In *FLAIRS Conference*, 764–769.

Mohammad, S. M., and Bravo-Marquez, F. 2017. Wassa-2017 shared task on emotion intensity. *arXiv preprint:1708.03700*.

Mohammad, S., and Turney, P. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*.

Mohammad, S. M.; Kiritchenko, S.; and Zhu, X. 2013. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint:1308.6242*.

Mohammad, S. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *ACL*.

Neviarouskaya, A.; Prendinger, H.; and Ishizuka, M. 2011. Affect analysis model: novel rule-based approach to affect sensing from text. *Natural Language Engineering*.

Niu, T., and Bansal, M. 2018. Polite dialogue generation without parallel data. *arXiv preprint:1805.03162*.

Palmer, M.; Gildea, D.; and Kingsbury, P. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics* 31(1):71–106.

Parikh, K. S.; Venkataram, V.; and Kalita, J. 2018. Towards a universal document encoder for authorship attribution.

Pearl, J. 2010. Causal inference. In *Causality: Objectives and Assessment*, 39–58.

Peng, R. D., and Hengartner, N. W. 2002. Quantitative analysis of literary styles. *The American Statistician*.

Peng, N.; Ghazvininejad, M.; May, J.; and Knight, K. 2018. Towards controllable story generation. In *Proceedings of the First Workshop on Storytelling*.

Peterson, K.; Hohensee, M.; and Xia, F. 2011. Email formality in the workplace: A case study on the enron corpus. In *Workshop on Languages in Social Media*.

Prabhumoye, S.; Tsvetkov, Y.; Salakhutdinov, R.; and Black, A. W. 2018. Style transfer through back-translation. *arXiv preprint:1804.09000*.

Preoţiuc-Pietro, D.; Schwartz, H. A.; Park, G.; Eichstaedt, J.; Kern, M.; Ungar, L.; and Shulman, E. 2016. Modelling valence and arousal in facebook posts. In *Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.

Sari, Y.; Stevenson, M.; and Vlachos, A. 2018. Topic or style? exploring the most useful features for authorship attribution. In *COLING*.

Sari, Y.; Vlachos, A.; and Stevenson, M. 2017. Continuous n-gram representations for authorship attribution. In *EACL*.

Semino, E., and Culpeper, J. 2002. *Cognitive stylistics: Language and cognition in text analysis*.

Serioussi, Y.; Smyth, R.; and Zukerman, I. 2011. Ghosts from the high court's past: Evidence from computational linguistics for dixon ghosting for mctiernan and rich. *UNSWLJ*.

Seroussi, Y.; Zukerman, I.; and Bohnert, F. 2010. Collaborative inference of sentiments from texts. In *International Conf on User Modeling, Adaptation, and Personalization*.

Seroussi, Y.; Zukerman, I.; and Bohnert, F. 2014. Authorship attribution with topic models. *Computational Linguistics*.

Sharma, A.; Vo, N.; Aditya, S.; and Baral, C. 2015. Identifying various kinds of event mentions in k-parser output. In *3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*.

Shen, T.; Lei, T.; Barzilay, R.; and Jaakkola, T. 2017. Style transfer from non-parallel text by cross-alignment. In *NIPS*.

Stamatatos, E. 2008. Author identification: Using text sampling to handle the class imbalance problem. *Information Processing & Management* 44(2):790–799.

Strapparava, C., and Mihalcea, R. 2007. Semeval-2007 task 14: Affective text. In *4th International Workshop on Semantic Evaluations*.

Strunk, W., and White, E. B. 1979. *The Elements of Style*.

Subasic, P., and Huettner, A. 2001. Affect analysis of text using fuzzy semantic typing. *IEEE Transactions on Fuzzy systems* 9(4):483–496.

Thurmair, G. 1990. Parsing for grammar and style checking. In *COLING*.

Tikhonov, A., and Yamshchikov, I. P. 2018. What is wrong with style transfer for texts? *arXiv preprint:1808.04365*.

Wang, Y., and Blei, D. M. 2018. The blessings of multiple causes. *arXiv preprint:1805.06826*.

Wiebe, J., and Mihalcea, R. 2006. Word sense and subjectivity. In *COLING and 44th Annual Meeting of the ACL*.

Wilson, A. 1979. *The Strange Ride of Rudyard Kipling: His Life and Works*. Viking Press.

Wise, K.; Bolls, P.; Myers, J.; and Sternadori, M. 2009. When words collide online: How writing style and video intensity affect cognitive processing of online news. *Journal of Broadcasting & Electronic Media* 53(4):532–546.