

# Discovering Stylistic Variations in Distributional Vector Space Models via Lexical Paraphrases

Xing Niu and Marine Carpuat

Department of Computer Science

University of Maryland, College Park

xingniu@cs.umd.edu, marine@cs.umd.edu

## Abstract

Detecting and analyzing stylistic variation in language is relevant to diverse Natural Language Processing applications. In this work, we investigate whether salient dimensions of style variations are embedded in standard distributional vector spaces of word meaning. We hypothesize that distances between embeddings of lexical paraphrases can help isolate style from meaning variations and help identify latent style dimensions. We conduct a qualitative analysis of latent style dimensions, and show the effectiveness of identified style subspaces on a lexical formality prediction task.

## 1 Introduction

Automatically analyzing and generating natural language requires capturing not only what is said, but also how it is said. Consider the sentences “he shot himself” and “he committed suicide”. The first one is less formal than the second one, and carries information beyond its literal meaning, such as the situation in which it might be used. Another example is “stamp show” vs. “philatelic exhibition”, English learners with limited vocabulary can use the former term since it is simpler.

As Natural Language Processing systems are deployed in a variety of settings, detecting and analyzing stylistic variations is becoming increasingly important, and is relevant to applications ranging from dialogue systems (Mairesse, 2008) to predicting power differences in social interactions (Danescu-Niculescu-Mizil et al., 2012).

In this work we aim to determine to what extent such stylistic variations are embedded in the topology of distributional vector space models. We focus on dense word embeddings, which pro-

vide a compact summary of word usage on the basis of the distributional hypothesis, and have been showed to capture semantic similarity and other lexical semantic relations (Mikolov et al., 2013; Baroni et al., 2014; Levy and Goldberg, 2014).

We hypothesize that differences between embeddings of words that share the same meaning are indicative of style differences. In order to test this hypothesis, we introduce a method based on Principal Component Analysis to identify salient dimensions of variations between word embeddings of lexical paraphrases.

Applying our method to word embeddings learned from two large corpora representing distinct genres, we conduct a qualitative analysis of the principal components discovered. It suggests that the principal components indeed discover variations that are relevant to style.

Second, we evaluate the style dimensions more directly, using them to distinguish more formal from less formal words. Formality is considered a key dimension of style variation (Heylighen and Dewaele, 1999), and it encompasses a range of finer-grained dimensions, including politeness, serious-trivial, etc (Irvine, 1979; Brown and Fraser, 1979).

The formality prediction task lets us evaluate empirically the impact of different factors in identifying style-relevant dimensions, including dimensionality of the subspace and the nature of the prediction method. We also conduct an error analysis revealing the limitation of predicting formality based on vector space models.

## 2 Background

Many studies of style variations have focused on the corpus or sentence level. For instance, multidimensional corpus analysis (Biber, 1995) relies on statistical analysis to identify the salient linguistic

co-occurrence patterns that underlie register variations. More recently, richer combinations of features have been used to predict style dimensions such as formality: (Pavlick and Tetreault, 2016) provide a thorough study of sentence-level formality and show that classifiers based on features including POS tags and dependency parses can predict formality as defined by the collective intuition of human annotators.

Here, we focus on identifying dimensions of style variations at the lexical level, motivated by the usefulness of word embeddings in many NLP tasks (Mikolov et al., 2013; Baroni et al., 2014), and by recent work that showed that meaningful ultradense subspaces that capture dimensions such as polarity and concreteness can be induced from word embeddings in a supervised fashion (Rothe and Schütze, 2016). Bolukbasi et al. (2016) induced a gender subspace using 10 human-selected gender pairs for reducing stereotypes. In contrast, we aim to discover style relevant dimensions without supervision, using instead lexical paraphrases to discover dimensions of variations that are not explained by semantic differences.

Prior work on evaluation of style factors at the word level has used standard word embeddings as features, and relied on external supervised methods to identify style relevant information in these embeddings. Brooke et al. (2010) proposed to score the formality of a word  $w$  by comparing its meaning to that of seed words of known formality using cosine similarity (Turney and Littman, 2003). Other approaches include work by Pavlick and Nenkova (2015) who used a unigram language model to capture the difference between lexical distributions across genres.

Beyond formality, analysis of stylistic variations from the point of view of the lexicon includes predicting term complexity, as annotated by non-native speakers (Paetzold and Specia, 2016). Preotiuc-Pietro et al. (2016) isolated stylistic differences associated with user attributes (gender, age) by using paraphrase pairs and word distributions similar to Pavlick and Nenkova (2015). Xu et al. (2012) used a machine translation model to paraphrasing Shakespeares plays into/from modern English.

### 3 Approach

Our approach to discovering stylistic variations in vector space models is based on the assumption

that these variations cannot be explained by differences in meaning, and they can be captured by salient dimensions of variation in the distributional spaces.

Lexical paraphrases should have the same meaning, and therefore their embeddings should be close to each other. When lexical paraphrases are not in the same location in the vector space, distances between them might be indicative of latent style variations. We discover such latent directions using Principal Component Analysis (PCA).<sup>1</sup>

Concretely, suppose  $e_i$  is the word embedding in the vector space for word  $w_i$ . Given pairs of word embeddings  $(e_1, e_2)$  for lexical paraphrases  $(w_1, w_2)$ , we subtracted them to get the relative direction  $d = e_1 - e_2$ .

For a given word pair, the difference vector might capture many things besides style variations. We hypothesize that the regularities among these differences for a large number of examples will reveal stylistic variations. Therefore, we then trained a PCA model on all directional vectors to get principal components  $(pc_k)$  capturing latent variations.

## 4 Qualitative Analysis of Latent Style Dimensions

### 4.1 Models Settings

The approach outlined above requires two types of inputs: (1) a word embedding space, and (2) a set of lexical paraphrases.

**Word Embeddings** We used word2vec (Mikolov et al., 2013) to build 300-dimensional vector space models for two corpora representing different genres. As suggested by Brooke et al. (2010), we selected the ICWSM 2009 Spinn3r dataset (English tier-1) as the training corpus (Burton et al., 2009). It consists of about 1.6 billion words in 7.5 million English blogs and is expected to have wide variety of language genres. We also compared it with the pre-trained 300-dimensional model of Google News<sup>2</sup>, which represents an even larger training corpus but in a narrower register. By working with two different

<sup>1</sup>Other algorithms for dimensionality reduction could also be leveraged to discover latent variations. E.g. multidimensional scaling (MDS) and t-distributed stochastic neighbor embedding (t-SNE).

<sup>2</sup><https://code.google.com/archive/p/word2vec/>

| $k$                             | Representative word pairs   |
|---------------------------------|---|
| <b>ICWSM 2009 Spinn3r Blogs</b> |   |
| 1                               | annulling • canceling    abolished • canceled    centre • center    emphasise • highlight<br>programme • program    imperatives • essentials    motorway • freeway    labour • labor<br>organised • organize    six-party • six-way    tranquility • serenity    tripartite • three-way                   |
| 2                               | spendings • expenditures    summons • subpoenas    anti-malaria • antimalarial<br>doctor • physician    falls • decreases    banned • prohibiting    fallen • decreased   |
| 3                               | decreased • receded    decreased • fallen    decreased • declined    decreased • shrank   |
| 4                               | agreements • understandings    unlimited • unbounded    disruptions • perturbations<br>discriminatory • discriminative    timetable • time-scale    amended • altered    ban • forbidden  |
| 5                               | underscored • underline    eliminated • delete    highlights • underline    widened • expand<br>widened • broaden    emphasises • underline    decreased • reduce    performed • fulfil   |
| 6                               | co-operate • collaborating    interdomain • cross-domain    cooperate • collaborating<br>origin • sourcing    executions • implementations    multifunctional • cross-functional  |
| 7                               | refusing • rebuffs    stopped • halts    stress • underlines    inspected • reviewed<br>withdrawals • withdraws    supervising • oversees    stress • emphasises    refused • rejects   |
| 8                               | restarting • revitalising    co-operation • collaborations    cooperation • collaborations<br>restart • resumes    cleric • clergymen    cooperates • collaborates    expel • expulsions  |
| 9                               | obtain • gain    multi-factor • multifactorial    restricts • hampers    retrieves • recovers<br>obstructs • hampers    revoking • canceling    contravened • breaches    invalidated • canceled  |
| 10                              | delete • eliminate    underline • stresses    underline • emphasises    schema • schemes<br>restarting • revitalising    decreased • reduce    underline • highlight    permissions • permits   |
| <b>Google News</b>              |   |
| 1                               | educator • educationist    ousts • deposes    exemptions • derogations    educator • educationalist<br>legal • juridical    truck • lorry    exceptions • derogations    accomplishments • attainments<br>roadway • carriageway    prohibit • proscribe    freeway • motorway    lucrative • remunerative |
| 2                               | standardize • standardizing    intercept • intercepting    evacuate • evacuating    isolate • isolating   |
| 3                               | destroys • demolishing    solves • resolving    impedes • obstructing    examines • investigating   |
| 4                               | falls • decreases    widens • increases    spends • expenditures    shrinks • decreases   |
| 5                               | infeasible • impracticable    impossible • impracticable    earmarks • allocates<br>unworkable • impracticable    confines • restricts    impractical • impracticable   |

Table 1: Representative word pairs for top principal components (indexed by  $k$ ) are listed for both blogs and news corpora. A mixed variation of formality and American-British English (grey-boxed) can be characterized by the first principal component, but the following principal components seem vaguer in terms of interpreting stylistic variations.

corpora, we aim to discover whether they share some common stylistic variations even though they have distinct word distributions.

**Lexical Paraphrases** PPDB 2.0 (Pavlick et al., 2015) provides automatically extracted lexical paraphrases with entailment annotations. We use the S-size pack and extracted word pairs with Equivalence entailment relation, which represent a cleaner subset of the original PPDB. This process yields 9427 paraphrase pairs found in the vocabulary of the blogs embeddings and 6988 pairs found in the vocabulary of the Google news

embeddings.

## 4.2 Analysis

We illustrate the principal components discovered in Table 1. For each of the  $k$ -th principal components, we can identify the most representative word pairs for that component by projecting all word pairs on  $pc_k$  and ranking pairs based on  $d \cdot pc_k$ .

The first observation is that the first principal components for both blogs and news corpora capture the pattern of American/British-English variations (grey-boxed in the Table). These might also

be related to the formality dimension of style, as British-English can be regarded to be more formal than American-English (Hurtig, 2006). However, not all representative word pairs fall in that category, and the nature of the variation between e.g., “annulling” and “canceling” is harder to characterize.

We can observe clues of stylistic variations in the subsequent (2nd+) principal components, but in general it is difficult to interpret each group. Several word pairs can be seen as illustrating formality variations (e.g., “falls” ↔ “decrease”, “delete” ↔ “eliminate”). Many word pairs are literally exchangeable but either one is preferred under certain context, such as “summons” vs. “subpoenas”, “decreased” vs. “fallen”, etc. Some principal components simply capture groups of words having semantic correlations, such as third PC of blogs and fourth PC of news (all contain “decrease/increase”), due to the biased word distribution of PPDB.

Although blogs and news corpora are expected to have different word distributions, they share the stylistic variation patterns mentioned above. One key difference between the principal components discovered in these two embedding spaces can be found in the second and third principal component of the news corpus, where “base (verb) ↔ present participle” is a dominant pattern, while it cannot be found in the top principal components of the blogs corpus.

Overall, this manual inspection suggests that the principal components do capture information that is relevant to style variations, even if they do not directly align to clear-cut style dimensions. Identifying how many top PCs are style-related (i.e. form a style subspace) is subjective and difficult. Therefore, we now turn to a quantitative evaluation.

## 5 Extrinsic Evaluation: Lexical Formality Scoring

We evaluate the usefulness of the latent dimensions discovered in Section 4 on a lexical formality prediction task. If the dimensions discovered are relevant to style, they should help predict formality with high accuracy.

## 5.1 Identifying A Style Subspace

### 5.1.1 Experimental Setup

**Task** Following Brooke et al. (2010), we used a list of 399 synonym pairs from a writing manual – *Choose the Right Word* (CTRW) (Hayakawa, 1994) – to evaluate the formality model. Given a pair of words, such as “hurry” vs. “expedite”, the task is to predict which is the more formal of the two.

**Ranking method** The predictions were made by linear SVM classifiers (similar to the method proposed by Brooke and Hirst (2014)). They were trained on 105 formal seed words and 138 informal seed words used by Brooke et al. (2010). Each word was represented by a feature vector in word2vec spaces or their subspaces. When ranking two words, we actually compared their distances to the separating hyperplane, i.e.  $w \cdot e - \rho$ , where  $w$ ,  $e$  and  $\rho$  are weight, embedding and bias.

**Embedding spaces** We first trained word2vec (W2V) models of blogs corpus with different space sizes (dimensionality=1-10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500). We then fixed the space size of word2vec models to 300 since it provides large enough original vector space and is a routinely used setting. All subspaces would be extracted from these 300-dimensional original spaces.

**Style subspaces** Next, we identified style subspaces (i.e. top PCs) using the PCA method introduced in Section 3. We examined every possible subspace size in the range of [1, 300] and denoted this method as PCA-PPDB.

For comparison, we also trained PCA subspaces using the seed words (PCA-seeds). Since seed words are not paraphrases, the PCA model was simply applied on word vectors. This method is based on the assumption that representative formal/informal words principally vary along the direction of formality.

### 5.1.2 Results

As illustrated in Figure 1, \*\*\* train indicates the training accuracy of SVM classifiers while \*\*\* test indicates the CTRW-pairs test accuracy.

The test accuracy of W2V curve has two peaks when dimensionality=10 (accuracy=0.798) and dimensionality=300 (accuracy=0.792). Considering the near-monotonicity of the training accuracy

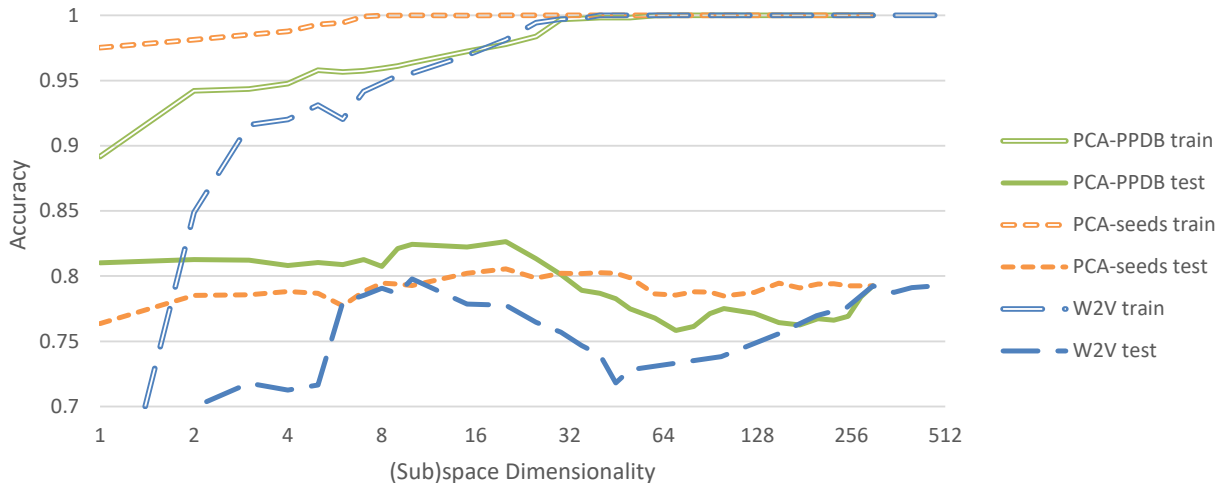


Figure 1: Train accuracy of formal/informal words classification and test accuracy of CTRW word-pair ranking v.s. the (sub)space dimensionality. An SVM-based formality model achieved the best test performance on subspaces identified by PCA on PPDB data.

curve, we attribute the trough around dimensionality=45 to over-fitting (increasing number of features) while attribute the rebound after that to more formality-related dimensions introduced.

Recall that we fixed the original spaces to 300 dimensions. The accuracy curve provides another reason to choose this number: 300-dimensional original spaces can model formality well by itself and the performance converges when  $\dim \geq 300$ .

Comparing `PCA-PPDB test` and `W2V test`, we can observe clear advantage of using subspaces that capture latent lexical variations. Even a single first principle dimension surpassed original word2vec models of any size, including the full 300-dimensional space which yielded a test accuracy of 0.792. Further improvements were achieved when 9th-21st principle dimensions were introduced (max accuracy=0.826) – back to Table 1, we can notice additional clues of formality variations from 9th PC.

The accuracy curves of `PCA-seeds` indicate that this model can fit the training set better with fewer dimensions than PPDB-based model but does not generalize as well to unseen test data. However, `PCA-seeds` still surpassed original word2vec models of any size.

## 5.2 SVM-based Ranking vs. Other Formality Models

We have discussed the effectiveness of modeling formality using a subspace of small size (1 for good results and  $\sim 20$  for best results). All analy-

ses so far were based on linear SVM, but can other sophisticated methods perform even better on the style-embedded subspaces?

### 5.2.1 Formality Models

We compare SVM with state-of-the-art lexical formality models based on vector space models, such as `SimDiff` (Brooke et al., 2010) and `DENSIFIER` (Rothe et al., 2016).

`SimDiff` (Brooke et al., 2010) scores the formality of a word  $w$  by comparing its meaning to that of seed words of known formality.<sup>3</sup> Intuitively,  $w$  is more likely formal if it is semantically closer to formal seed words than to informal seed words. Formally, given a formal word set  $S_f$  and an informal word set  $S_i$ , `SimDiff` scores a word  $w$  by

$$\text{score}(w) = \frac{1}{|S_f|} \sum_{v \in S_f} e_w \cdot e_v - \frac{1}{|S_i|} \sum_{v \in S_i} e_w \cdot e_v$$

Further manipulations such as score de-biasing and normalization were also introduced in (Brooke et al., 2010), but they would not affect rankings examined by our evaluation.

`DENSIFIER` (Rothe et al., 2016) is a supervised learning algorithm that transforms word embeddings into pre-defined ultra-dense orthogonal dimensions such as sentiment and concreteness. Under the formality ranking scenario, it optimizes a

<sup>3</sup>While Brooke et al. (2010) used cosine to measure the similarity in LSA spaces, we found that dot product yields better results with word2vec embeddings.

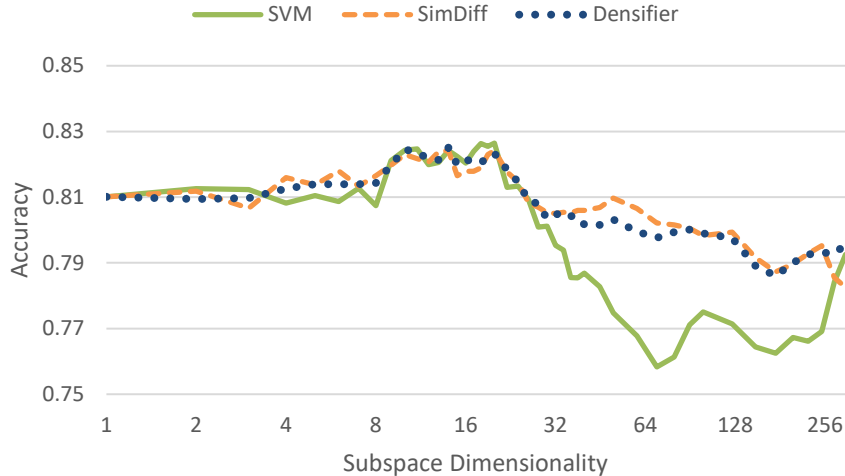


Figure 2: Test accuracy of CTRW word-pair ranking v.s. the subspace dimensionality. All formality models achieved similar performance on subspaces of size 9-21 identified by PCA-PPDB.

| Incorrect Examples |               |       |        |             | Correct Examples |             |        |       |             |
|--------------------|---------------|-------|--------|-------------|------------------|-------------|--------|-------|-------------|
| $w_1$              | $w_2$         | $s_1$ | $s_2$  | $s_2 - s_1$ | $w_1$            | $w_2$       | $s_1$  | $s_2$ | $s_2 - s_1$ |
| crony              | friend ††     | 0.667 | -1.414 | -2.081      | grill †          | interrogate | -1.370 | 1.212 | 2.581       |
| conceit            | vanity †      | 1.107 | -0.697 | -1.804      | excuse †         | remit       | -0.608 | 2.001 | 2.609       |
| present †          | gift          | 1.017 | -0.732 | -1.749      | gardening ††     | tillage     | -0.846 | 1.795 | 2.641       |
| shiv               | knife †       | 0.681 | -0.863 | -1.543      | get ††           | obtain      | -1.435 | 1.296 | 2.731       |
| quotation          | quote †       | 0.910 | -0.594 | -1.504      | hurry †          | expedite    | -1.632 | 1.174 | 2.806       |
| frighten           | scare †       | 0.157 | -1.244 | -1.400      | catch ††         | apprehend   | -1.443 | 1.381 | 2.824       |
| phony              | fake †        | 0.237 | -1.100 | -1.337      | watch †          | observe     | -1.628 | 1.264 | 2.892       |
| parched            | dehydrated †  | 0.173 | -1.035 | -1.209      | loud ††          | clamorous   | -1.304 | 1.819 | 3.123       |
| punish †           | chase         | 0.260 | -0.697 | -0.956      | quote ††         | adduce      | -0.594 | 2.529 | 3.123       |
| penetrating †      | perspicacious | 1.527 | 0.644  | -0.883      | beach ††         | littoral    | -1.116 | 2.143 | 3.259       |

Table 2: Top (mis-)predicted CTRW word pairs, where  $s_i$  is the SVM (formality) score for word  $w_i$ .  $w_2$  is supposed to be more formal than  $w_1$ . † This word is more frequent than the other in a pair according to the blogs corpus. (†/ ††/ ††† means at least 10/100/1000 times more.)

formality dimension (transition vector) that aims at separating words in  $S_f$  and words in  $S_i$ , and grouping words in the same set.

### 5.2.2 Results

All three formality scoring models (i.e. linear SVM, SimDiff and DENSIFIER) were applied to subspaces extracted from 300-dimensional word2vec spaces using PCA on PPDB data. Figure 2 shows that three models achieves nearly identical accuracy on subspaces with size smaller than 28.<sup>4</sup> Furthermore, we also compared the formality directions discovered by linear SVM (coefficient  $w$ ) and Densifier (transition vector). For any dimensionality, the cosine similarity between them are larger than 0.8. It is even larger than 0.9

<sup>4</sup>SVM could also have similar accuracy curve after dimension=28 if an RBF kernel was used.

when  $\dim \geq 21$ . These suggest that the choice of ranking models has marginal impact, therefore identifying the style subspace plays a more critical role in modeling formality.

### 5.3 Error Analysis

Identified subspaces capture formality decently in terms of ranking lexical formality – as high as 0.826 accuracy in the CTRW dataset (based on the best performing model, i.e. a linear SVM trained on a 20-dimensional subspace identified by PCA-PPDB). The question then arises: what types of errors contribute to the incorrect predictions?

Top (mis-)predicted CTRW word pairs are listed in Table 2, where  $s_i$  is the SVM (formality) score for word  $w_i$ .  $w_2$  is supposed to be more formal than  $w_1$ .

One category of errors roots in the mechanism

of vector space models such as word2vec: they are all based on word co-occurrence patterns, which sometimes introduce unwanted biases. For example, “crony” itself is an informal synonym of “friend” in our dataset. However, “crony capitalism” is a tightly glued economy term. For comparison, the formality score of “capitalism” is 0.966, which is very close to 0.667 of “crony”.

Ambiguity is another key factor that influences the formality scoring based on vector space models. Arora et al. (2016) pointed out that in the vector space, a word having multiple meanings lies in middle of its senses. Consequently, its formality score is also controlled by all its senses. We can find many ambiguous words in the list of incorrect examples, such as “vanity” (clothing store, singer), “present”, “shiv” (Hindu god), “parched” (film), “chasen” (surname, band), etc.

Last but not least, word frequency is a strong signal of predicting formality, but predictions can easily be stereotyped. We used word frequencies in the blogs corpus to rank CTRW word pairs and got an accuracy as high as 0.771 (by arguably treating more frequent as less formal). Projecting to the top (in)correct examples, a † symbol is placed behind the more frequent word in a pair. We can observe that top correctly ranked pairs followed the more-frequent-less-formal rule. However, this rule also biased the prediction to some incorrectly ranked pairs. Frequency information is not designed to be embedded into Word2vec models, but it still can be partially reconstructed (Rothe et al., 2016).

In a nutshell, formality models based on vector space models suffers from the limitation that a word representation is affected by word association, word sense and word frequency.

## 6 Conclusion

We presented an approach to discovering stylistic variations in distributional vector spaces using lexical paraphrases. Qualitative analysis suggests that the principle components discovered by PCA indeed capture variations related to style. Evaluation on a formality prediction task demonstrates the benefits of the induced subspace to detect style variations. We also compared the impact of different factors in identifying style-relevant dimensions such as the training data for PCA, the dimensionality of subspaces and the nature of prediction methods. Finally, the error analysis indi-

cated some intrinsic limitation of comparing style (formality) based on vector space models.

## References

- Sanjeev Arora, Yuanzhi Li, Yingyu Liang, Tengyu Ma, and Andrej Risteski. 2016. Linear algebraic structure of word senses, with applications to polysemy. *CoRR*, abs/1601.03764.
- Marco Baroni, Georgiana Dinu, and Germán Kruszewski. 2014. Don’t count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *ACL (1)*, pages 238–247. The Association for Computer Linguistics.
- Douglas Biber. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge University Press.
- Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam Tauman Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *NIPS*, pages 4349–4357.
- Julian Brooke and Graeme Hirst. 2014. Supervised ranking of co-occurrence profiles for acquisition of continuous lexical attributes. In *COLING*, pages 2172–2183. ACL.
- Julian Brooke, Tong Wang, and Graeme Hirst. 2010. Automatic acquisition of lexical formality. In *COLING (Posters)*, pages 90–98. Chinese Information Processing Society of China.
- Penelope Brown and Colin Fraser. 1979. Speech as a marker of situation. In *Social Markers in Speech*, pages 33–62. Cambridge University Press.
- Kevin Burton, Akshay Java, and Ian Soboroff. 2009. *The ICWSM 2009 Spinn3r dataset*. In *Proceedings of the Third Annual Conference on Weblogs and Social Media (ICWSM 2009)*, San Jose, CA.
- Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon M. Kleinberg. 2012. Echoes of power: language effects and power differences in social interaction. In *WWW*, pages 699–708. ACM.
- Samuel Ichiye Hayakawa. 1994. *Choose the right word*. Collins Reference.
- Francis Heylighen and Jean-Marc Dewaele. 1999. Formality of language: Definition, measurement and behavioral determinants. *Interner Bericht, Center “Leo Apostel”, Vrije Universiteit Brussel*.
- Markus Hurtig. 2006. *Varieties of English in the Swedish classroom*. Karlstad University: Unpublished C-Essay.
- Judith T. Irvine. 1979. Formality and informality in communicative events. *American Anthropologist*, 81(4):773–790.

- Omer Levy and Yoav Goldberg. 2014. Linguistic regularities in sparse and explicit word representations. In *CoNLL*, pages 171–180. ACL.
- François Mairesse. 2008. *Learning to Adapt in Dialogue Systems: Data-Driven Models for Personality Recognition and Generation*. Ph.D. thesis, University of Sheffield, United Kingdom.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.
- Gustavo Paetzold and Lucia Specia. 2016. Semeval 2016 task 11: Complex word identification. In *SemEval@NAACL-HLT*, pages 560–569. The Association for Computer Linguistics.
- Ellie Pavlick and Ani Nenkova. 2015. Inducing lexical style properties for paraphrase and genre differentiation. In *HLT-NAACL*, pages 218–224. The Association for Computational Linguistics.
- Ellie Pavlick, Pushpendre Rastogi, Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. 2015. PPDB 2.0: Better paraphrase ranking, fine-grained entailment relations, word embeddings, and style classification. In *ACL (2)*, pages 425–430. The Association for Computer Linguistics.
- Ellie Pavlick and Joel R. Tetreault. 2016. An empirical analysis of formality in online communication. *TACL*, 4:61–74.
- Daniel Preotiuc-Pietro, Wei Xu, and Lyle H. Ungar. 2016. Discovering user attribute stylistic differences via paraphrasing. In *AAAI*, pages 3030–3037. AAAI Press.
- Sascha Rothe, Sebastian Ebert, and Hinrich Schütze. 2016. Ultradense word embeddings by orthogonal transformation. In *HLT-NAACL*, pages 767–777. The Association for Computational Linguistics.
- Sascha Rothe and Hinrich Schütze. 2016. Word embedding calculus in meaningful ultradense subspaces. In *ACL (2)*. The Association for Computer Linguistics.
- Peter D. Turney and Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Trans. Inf. Syst.*, 21(4):315–346.
- Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, and Colin Cherry. 2012. Paraphrasing for style. In *COLING*, pages 2899–2914. Indian Institute of Technology Bombay.