Understanding and Quantifying Creativity in Lexical Composition

Polina Kuznetsova **Jianfu Chen** **Yejin Choi**
Department of Computer Science
Stony Brook University
Stony Brook, NY 11794-4400
{pkuznetsova,jianchen,ychoi}@cs.stonybrook.edu

Abstract

Why do certain combinations of words such as *"disadvantageous peace"* or *"metal to the petal"* appeal to our minds as interesting expressions with a sense of creativity, while other phrases such as *"quiet teenager"*, or *"geometrical base"* not as much? We present statistical explorations to understand the characteristics of lexical compositions that give rise to the perception of being original, interesting, and at times even artistic. We first examine various correlates of perceived creativity based on information theoretic measures and the connotation of words, then present experiments based on supervised learning that give us further insights on how different aspects of lexical composition collectively contribute to the perceived creativity.

1 Introduction

An essential property of natural language is the generative capacity that makes it possible for people to express indefinitely many thoughts through indefinitely many different ways of composing phrases and sentences (Chomsky, 1965). The possibility of novel, creative expressions never seems to exhaust. Various types of writers, such as novelists, journalists, movie script writers, and creatives in advertising, continue creating novel phrases and expressions that are original while befitting in expressing the desired meaning in the given situation. Consider unique phrases such as *"geological split personality"*, or *"intoxicating Shangri-La of shoes"*,[1] that continue flowing into the online text drawing attention from readers.

Writers put significant effort in choosing the perfect words in completing their compositions, as a well-chosen combination of words is impactful in readers' minds for rendering the precise intended meaning, as well as stimulating an increased level of cognitive responses and attention. Metaphors in particular, one of the quintessential forms of linguistic creativity, have been discussed extensively by studies across multiple disciplines, e.g., Cognitive Science, Psychology, Linguistics, and Literature (e.g., Lakoff and Johnson (1980), McCurry and Hayes (1992), Goatly (1997)). Moreover, recent studies based on fMRI begin to discover biological evidences that support the impact of creative phrases on people's minds. These studies report that unconventional metaphoric expressions elicit significantly increased involvement of brain processing when compared against the effect of conventional metaphors or literal expressions (e.g., Mashal et al. (2007), Mashal et al. (2009)).

Several linguistic elements, e.g., syntax, semantics, and pragmatics, are likely to be working together in order to lead to the perception of creativity. However, their underlying mechanisms by and large are yet to be investigated. In this paper, as a small step toward quantitative understanding of linguistic creativity, we present a focused study on lexical composition two content words.

Being creative, by definition, implies qualities such as being unique, novel, unfamiliar or unconventional. But not every unfamiliar combination of words would appeal as creative. For example, unfa-

[1] Examples from New York Times articles in 2013.

Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1246–1258,
Seattle, Washington, USA, 18-21 October 2013.

miliar biomedical terms, e.g., *"cardiac glycosides"*, are only informative without appreciable creativity. Similarly, less frequent combinations of words, e.g., *"rotten detergent"* or *"quiet teenager"*, though describing situations that are certainly uncommon, do not bring about the sense of creativity. Finally, some unique combinations of words can be just nonsensical , e.g., *"elegant glycosides"*.

Different studies assumed different definitions of linguistic creativity depending on their context and end goals (e.g., Chomsky (1976), Zhu et al. (2009), Gervás (2010), Maybin and Swann (2007), Carter and McCarthy (2004)). In this paper, as an operational definition, we consider a phrase creative if it is (a) unconventional or uncommon, and (b) expressive in an interesting, imaginative, or inspirational way.

A system that can recognize creative expressions could be of practical use for many aspiring writers who are often in need of inspirational help in searching for the optimal choice of words. Such a system can also be integrated into automatic assessment of writing styles and quality, and utilized to automatically construct a collection of interesting expressions from the web, which may be potentially useful for enriching natural language generation systems.

With these practical goals in mind, we aim to understand phrases with linguistic creativity in a broad scope. Similarly as the work of Zhu et al. (2009), our study encompasses phrases that evoke the sense of interestingness and creativity in readers' minds, rather than focusing exclusively on clearly but narrowly defined figure of speeches such as *metaphors* (e.g., Shutova (2010)), *similes* (e.g., Veale et al. (2008), Hao and Veale (2010)), and *humors* (e.g., Mihalcea and Strapparava (2005), Purandare and Litman (2006)). Unlike the study of Zhu et al. (2009), however, we concentrate specifically on how combinations of different words give rise to the sense of creativity, as this is an angle that has not been directly studied before. We leave the roles of syntactic elements as future research.

We first examine various correlates of perceived creativity based on information theoretic measures and the connotation of words, then present experiments based on supervised learning that give us further insights on how different aspects of lexical composition collectively contribute to the perceived creativity.

2 Theories of Creativity and Hypotheses

Many researchers, from the ancient philosophers to the modern time scientists, have proposed theories that attempt to explain the mechanism of creative process. In this section, we draw connections from some of these theories developed for general human creativity to the problem of quantitatively interpreting linguistic creativity in lexical composition.

2.1 Divergent Thinking and Composition

Divergent thinking (e.g., McCrae (1987)), which seeks to generate multiple unstereotypical solutions to an open ended problem has been considered as the key element in creative process, which contrasts with *convergent* thinking that find a single, correct solution (e.g., Cropley (2006)). Applying the same high-level idea to lexical composition, divergent composition that explores an unusual, unconventional set of words is more likely to be creative.

Note that the key novelty then lies in the *compositional* operation itself, i.e., the act of putting together a set of words in an unexpected way, rather than the rareness of individual words being used. In recent years there has been a swell of work on compositional distributional semantics that captures the compositional aspects of language understanding, such as sentiment analysis (e.g., Yessenalina and Cardie (2011), Socher et al. (2011)) and language modeling (e.g., Mitchell and Lapata (2009), Baroni and Zamparelli (2010), Guevara (2011), Clarke (2012), Rudolph and Giesbrecht (2010)). However, none has examined the compositional nature in quantifying creativity in lexical composition.

We consider two computational approaches to capture the notion of creative composition. The first is via various information theoretic measures, e.g., relative entropy reduction, to measure the surprisal of seeing the next word given the previous word. The second is via supervised learning, where we explore different modeling techniques to capture the statistical regularities in creative compositional operations. In particular, we will explore (1) compositional operations of vector space models, (2) kernels capturing the non-linear composition of different dimensions in the meaning space, (3) the use of neural networks as an alternative to incorporate non-linearity in vector composition. (See §5).

2.2 Latent Memory and Creative Semantic Subspace

Although we expect that unconventional composition has a connection to creativeness of resulting phrases, that alone does not explain many counter examples where the composition itself is uncommon but the resulting expression is not creative due to lack of interestingness or imagination, e.g., *"room and water"*.[2] Therefore, we must consider additional conditions that give rise to creative phrases.

Let $\mathcal{S}$ represent the semantic space, i.e., the set of all possible semantic representation that can be expressed by a phrase that is composed of two content words.[3] Then we hypothesize that some subsets of semantic space $\{\mathcal{S}_i | \mathcal{S}_i \subset \mathcal{S}\}$ are semantically futile regions for appreciable linguistic creativity, regardless of how novel the composition in itself might be. Such regions may include technical domains such as law or pharmacology. Similarly, we expect semantically fruitful subsets of semantic space where creative expressions are more frequently found. For instance, phrases such as *"guns and roses"* and *"metal to the petal"* are semantically close to each other and yet both can be considered as interesting and creative (as opposed to one of them losing the sense of creativity due to its semantic proximity to the other).

This notion of creative semantic subspace connects to theories that suggest that latent memories serve as motives for creative ideas and that one's creativity is largely depending on prior experience and knowledge one has been exposed to (e.g., Freud (1908), Necka (1999), Glaskin (2011), Cohen and Levinthal (1990), Amabile (1997)), a point also made by Einstein: *"The secret to creativity is knowing how to hide your sources."*

Figure 5 presents visualized supports for creative semantic subspace,[4] where we observe that phrases in the neighborhood of legal terms are generally not creative, while the semantic neighborhood of *"kingdom"* and *"power"* is relatively more fruitful for composing *creative* (i.e., unique and uncommon while being imaginative and interesting, per our operational definition of creativity given in §1) word pairs, e.g., *invisible empire"*. In our empirical investigation, this notion of semantically fruitful and futile semantic subspaces are captured using distributional semantic space models under supervised learning framework (§5).

[2]With additional context this example may turn into a creative one, but for simplicity we focus on phrases with two content words considered out of context.

[3]Investigation on recursive composition of more than two content words and the influence of syntactic packaging is left as future research.

[4]See §6 for more detailed discussion.

Source	# of uniq words	# of sent	Avg sent len	Entropy
QUOTES^{raw}	29498	49402	28	173.05
GLOSSES^{raw}	20869	7745	53	96.79

Table 1: Entropy of word distribution in datasets

Dataset	# of word pairs			percentage
	total	#(-)	#(+)	#(+)/total %
GLOSSES	1912	149	18	0.94
QUOTES	3298	204	35	1.06

Table 2: Distribution of creative(+)/common(-) word pairs over **GLOSSES** and **QUOTES** dataset.

2.3 Affective Language

Another angle we probe is the connection between creative expressions and the use of affective language. This idea is supported in part by previous research that explored the connection between figurative languages such as metaphors and sentiment (e.g., Fussell and Moss (1998), Rumbell et al. (2008), Rentoumi et al. (2012)). The focus of previous work was either on interpretation of the sentiment in metaphors, or the use of metaphors in the description of affect. In contrast, we aim to quantify the correlation between creative expressions (beyond metaphors) and the use of sentiment-laden words in a more systematic way. This exploration has a connection to the creative semantic subspace discussed earlier (§2.2), but pays a more direct attention to the aspect of sentiment and connotation.

3 Creative Language Dataset

We start our investigation by considering two types of naturally existing collection of sentences: (1) quotes and (2) dictionary glosses. We expect that quotes are likely to be rich in creative expressions, while dictionary glosses stand in the opposite spec-

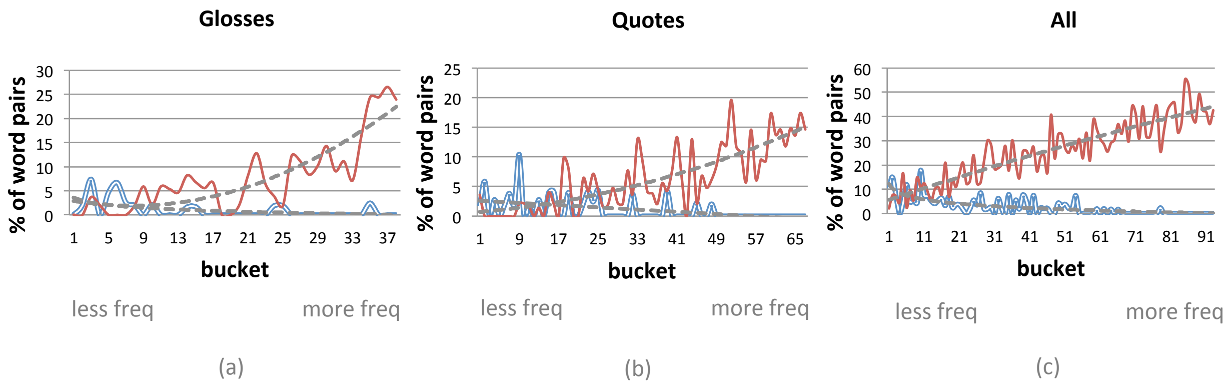

Figure 1: Distribution of *creative* (double lines in blue) versus *common* (single lines in red) word pairs with varying ranges of frequencies (x-axis) for **GLOSSES**, **QUOTES** and both datasets combined.

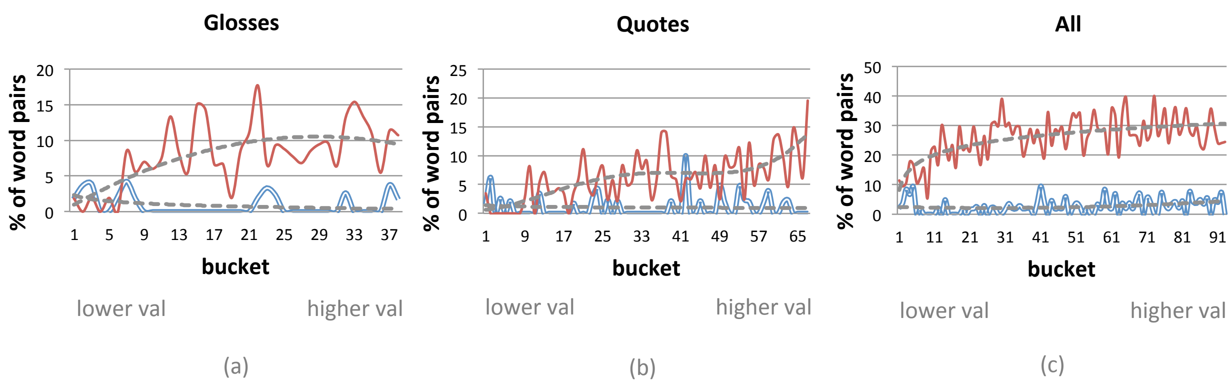

Figure 2: Distribution of *creative* (double lines in blue) versus *common* (single lines in red) word pairs with varying ranges of PMI values (x-axis) for **GLOSSES**, **QUOTES** and both datasets combined.

trum of being creative.

QUOTES^{raw}: We crawled inspirational quotes from "Brainy Quote".[5]

GLOSSES^{raw}: We collected glosses from Oxford Dictionary and Merriam-Webster Dictionary.[6] Overall we crawled about 8K definitions. Table 1 shows statistics of the dataset.[7]

Entropy of word distribution We conjecture that QUOTES and GLOSSES are different in terms of word variety, which can be quantified by the entropy of word distributions. To compute the entropy for each dataset, we use ngram statistics from the corresponding dataset to measure the probability of each word. As expected, QUOTES dataset has higher entropy than GLOSSES in Table 1.

3.1 Creative Word Pairs

We extract word pairs corresponding to the following syntactic patterns: [NN NN], [JJ NN], [NN JJ] and [JJ JJ]. Not all pairs from QUOTES^{raw} are creative, and likewise, not all pairs from GLOSSES^{raw} are uncreative. Therefore, we perform manual annotations to a subset of the collected pairs as follows. We obtain a small subset of pairs by applying stratified sampling based on bigram frequency buckets: first we sort word pairs by their bigram frequencies obtained from Web 1T corpus (Brants and Franz (2006)), group them into consecutive fre-

[5] http://www.brainyquote.com/

[6] http://oxforddictionaries.com/ and http://www.merriam-webster.com/. We only consider words appearing in both dictionaries to avoid unusual words such as compound words, e.g., "zero-base".

[7] QUOTES^{raw} contain 30K unique words and GLOSSES^{raw} has 20K unique words. QUOTES^{raw} have much arger number of sentences, while its average sentence is shorter.

quency buckets each of which containing 400 word pairs, then sample 40 word pairs from each bucket.

We label word pairs using Amazon Mechnical Turk (AMT) (e.g., Snow et al. (2008)). We ask three turkers to score each pair in 1-5 scale, where 1 is the least creative and 5 is the most creative. We then obtain the final creativity scale score by averaging the scores over 3 users. In addition, we ask turkers a series of yes/no questions to help turkers to determine whether the given pair is creative or not.[8] We determine the final label of a word pair based on two scores, creativity scale score and yes/no question-based score. If creativity scale score is 4 or 5 and question-based score is positive, we label the pair as creative. Similarly, if creativity scale score is 1 or 2 and question-based score is negative, we label the pair as common. We discard the rest from the final dataset. This filtering process is akin to the removal of neural sentiment in the early work of sentiment analysis (e.g., Pang et al. (2002)).[9] Table 2 shows the statistics of the resulting dataset.

Creative Pairs and their Frequencies: To gain insights on the stratified sample of word pairs, we plot the label ($\in \{creative, common\}$) distribution of word pairs as a function of simple statistics, such as a range (bucket) of bigram frequencies or PMI values of the given pair of words. Both bigram frequencies and PMI scores are computed based on Google Web 1T corpus Brants and Franz (2006). Figure 1 shows the results for word frequencies. As expected, word pairs with high frequencies are much more likely to be common, while word pairs with low frequencies can be either of the two. Also as expected, pairs extracted from QUOTES are relatively more likely to be creative than those from GLOSSES. In any case, it is clear that not all rare pairs are creative.

Creative Pairs and their PMI Scores: Similarly as above, Figure 2 plots the relation between the distribution of labels of word pairs and their corresponding PMI. As expected, pairs with high PMI are more likely to be common, though the trend is not as skewed as before.

Common	Creative
quiet teenager	inglorious success
constant longitude	thorny existence
watery juice	relaxed symmetry
noble political	sardonic destiny
diet cooking	dispassionate history
verbal interpretation	poetical enthusiasm
unwelcome situation	verbal beauty
migratory tuna	earth breathe
lousy businessman	disadvantageous peace
terrific marriage	alchemical marriage
solved issue	deep nonsense

Table 3: Sample Creative / Common Word Pairs

Final Dataset: From our initial annotation study, it became apparent to us that creative pairs are very rare, perhaps not surprisingly, even among infrequent pairs. In order to build the word pair corpus with as many creative pairs as possible, we focus on infrequent word pairs for further annotation, from which we construct a larger and balanced set of creative and common word pairs, with 394 word pairs for each class. The specific construction procedure is as follows: first combine all of the word pairs extracted from both QUOTES^{raw} and GLOSSES^{raw} as a single dataset, sort them by bigram frequency, group them into consecutive frequency buckets each of which has 40 word pairs; finally balance each frequency bucket, by discarding word pairs with higher frequency value from the larger class in that bucket. Examples of labeled word pairs are shown in Table 3. Hereafter we use this balanced dataset of word pairs for all experiments.[10]

4 Creativity Measures

4.1 Information Measures

In this section we explore information theoretic measures to quantify the *surprisal* aspect of creative word pairs, relating to the divergent, compositional nature of creativity discussed in §2.1.

Entropy of Context Seeing a word w changes our expectation on what might follow next. Some words have stronger selective preference (higher entropy) than others.

[8] E.g., *"is this word combination boring and not original?"* or *"does it provoke unusual imagination?"*.

[9] Cohen's Kappa and Pearson Correlation on the filtered data are 0.69 and 0.72 respectively. Corresponding scores for the unfiltered data drop to 0.26 and 0.29 respectively. All the experiments are performed on the filtered data.

[10] The resulting dataset is available at `http://www.cs.stonybrook.edu/~pkuznetsova/creativity/`

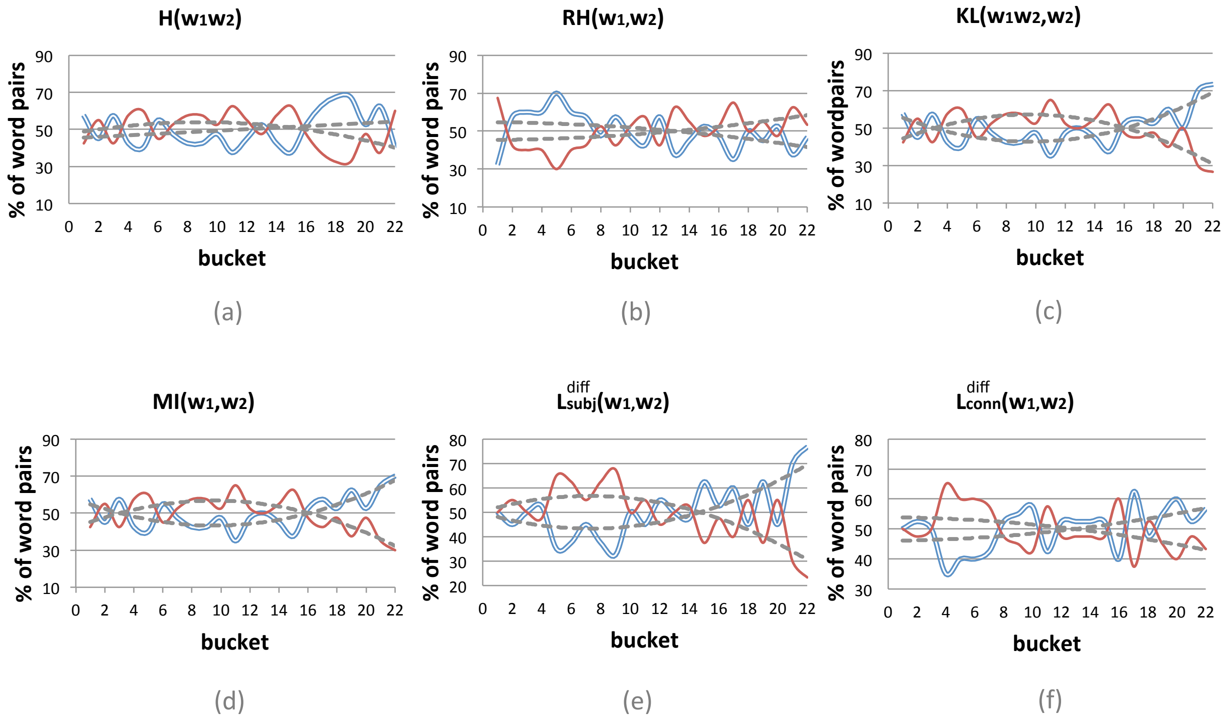

Figure 3: Distribution of *creative* (double lines in blue) versus *common* (single lines in red) word pairs with varying ranges of information or polarity measures (x-axis).

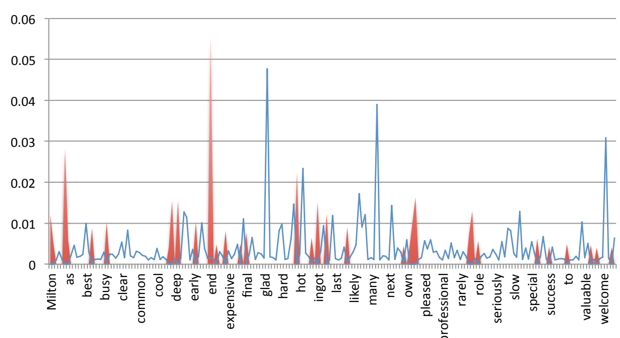

Figure 4: Conditional probability of neighboring words for "inglorious" (filled / red) and "very" (unfilled / blue).

For instance, the entropy after seeing *"very"* would be higher than that after seeing *"inglorious"*, as the former can be used in a wider variety of context than the later. Figure 4 visualizes relatively more skewed distribution of *"inglorious"*. We compute the entropy of future context conditioning on w_1, w_2 and w_1w_2, which we denote as $H(w_1)$, $H(w_2)$, $H(w_1w_2)$ respectively, latter is shown in Figure 3 – a.[11]

Relative Entropy Transformation In order to focus more directly on the relative change of entropy as a result of composition, we compute Relative Entropy Transformation:

$$RH(w_1, w_2) = \frac{|H(w_1) - H(w_1w_2)|}{H(w_1) + H(w_1w_2)} \tag{1}$$

As expected (Figure 3 – b and Table 4), this relative quantity captures creativity better than the absolute measure $H(w_1w_2)$ computed above. The idea behind this measure has a connection to uncertainty reduction in psycholinguistic literature (e.g., Frank (2010), Hale (2003), Hale (2006)).

KL divergence To capture unusual combinations of words, we compare the difference between the distributional contexts of w_1 and w_1w_2 so that

$$KL(w_1w_2, w_1) = \sum_{w_i \in V} P(w_i|w_1, w_2) \log \frac{P(w_i|w_1, w_2)}{P(w_i|w_1)} \tag{2}$$

Figure (3 – c) shows that $KL(w_1w_2, w_1)$[12] is among

[11] As before, language models are drawn from Google Web 1T corpus Brants and Franz (2006).

[12] We also compute $KL(w_1, w_2)$ in a similar manner as $KL(w_1w_2, w_1)$

the effective measures in capturing creative pairs.

Mutual Information Finally, we consider mutual information (Figure 3 – d):

$$MI(w_1, w_2) = \sum_{w_i \in V} P(w_i|w_1, w_2) \times \log \frac{P(w_i|w_1, w_2)}{P(w_i|w_1) \cdot P(w_i|w_2)} \quad (3)$$

Correlation coefficients Pearson coefficients for all measures are shown in Table 4. Interestingly, information theoretic measures that compare the distribution of word's context, such as $RH(w_1, w_2)$, $KL(w_1w_2, w_1)$ and $MI(w_1, w_2)$, capture the surprisal aspect of creativity better than simple frequencies or PMI scores that do not consider contextual changes. But even for those cases when the correlation is statistically significant, the values are not too high. We conjecture that there are two reasons for this. First, Pearson assumes linear correlations, hence not sensitive enough to capture non-linear correlations that are evident in graphs shown in Figure 3. Second, these measures only capture the surprisal aspect of creativity, missing the other important qualities: interestingness or imaginativeness.

4.2 Sentiment and Connotation

Next we investigate the connection between creativity and sentiment, as illustrated in §2.3. We consider both sentiment (more explicit) and connotation (more implicit) words,[13] and consider them with or without distinguishing the polarity (i.e., positive, negative). To determine sentiment and connotation, we use lexicons provided by OpinionFinder (Wilson et al. (2005)) and Feng et al. (2013) respectively. We denote polarity of a word w_i as $L(w_i)$.[14] When w_i has a negative polarity $L(w_i)$ is assigned a value of -1, and when w_i is positive $L(w_i)$ is equal to 1. We assume that a word is neutral when it is not in the lexicon, assigning 0 to $L(w_i)$. For a word pair w_1w_2 we compute absolute difference $L^{diff}(w_1, w_2)$ between polarities of tokens in a word pair in order to catch examples such as *"inglorious success"*.

[13]E.g., expressions such as *"blue sky"* or *"white sand"* are not sentiment-laden, but do have positive connotation.

[14]We denote polarity from OpinionFinder as L_{subj} and connotation as L_{conn}

Measure	Corr Coeff	p-value*	adj p-value**
pointwise, noncontextual			
$Freq(w_1w_2)$	0.014	0.67	0.86
$PMI(w_1, w_2)$	0.011	0.75	0.86
information theoretic, contextual			
$E(w_1)$	-0.038	0.26	0.49
$E(w_2)$	**-0.126**	0.00019	0.00083
$E(w_1, w_2)$	0.013	0.71	0.86
$RH(w_1, w_2)$	**0.113**	0.00081	0.0024
$KL(w_1w_2, w_1)$	**0.134**	7.152-05	0.00054
$KL(w_1, w_2)$	**-0.080**	0.018	0.039
$MI(w_1, w_2)$	**0.125**	0.00022	0.00083
sentiment & connotation			
$L_{subj}(w_1)$	0.006	0.87	0.87
$L_{subj}(w_2)$	0.031	0.36	0.60
$L_{subj}^{diff}(w_1, w_2)$	**0.168**	6.67e-07	1.00e-05
$L_{conn}(w_1)$	0.023	0.49	0.74
$L_{conn}(w_2)$	0.008	0.80	0.86
$L_{conn}^{diff}(w_1, w_2)$	**0.082**	0.015	0.038

Table 4: Pearson correlation between various measures and creativity of word pairs. Boldface denotes statistical significance ($p \leq 0.05$).
note *: Two-tailed p-value, 394 word pairs per class
note **: We used Benjamini-Hochberg method to adjust p-values for multiple tests

Table 4 shows Pearson coefficient for sentiment and connotation based measures. It turns out that polarity of each word on its own does not have a high impact on the creativity of a word pair. Rather, it is the difference between the two words that gives rise the sense of creativity.

4.3 Learning to Recognize Creativity

Now we put together all measures explored in §4.1 and 4.2 in a supervised-learning framework. As expected, rather than either one alone, the combination of various measures leads to the best performance:

$$\begin{aligned} \vec{F}_{12} = {} & [RH(w_1, w_2); KL(w_1, w_2); H(w_1w_2); \\ & L_{conn}^{diff}(w_1, w_2); PMI(w1, w_2); \\ & H(w_2); KL(w_1w_2, w_1); KL(w_2, w_1); \\ & L_{subj}^{diff}(w_1, w_2); MI(w_1, w_2); \\ & Freq(w_1w_2); H(w_1)] \end{aligned}$$

Table 5 shows the performance of the above feature vector with 12 features using libsvm (Chang and Lin, 2011). We use C-Support Vector Classification (C-SVC). Performance is reported in accuracy using 5-fold cross validation.[15]

[15]Among these 12 features, the feature selection algorithm

5 Learning Creative Pairs with Distributional Semantic Vectors

The measures explored in §4 were largely uninformed of distributional semantic dimensions of each word. However, in order to pursue the conceptual aspect of creativity illustrated in §2.2, that is, the notion of semantic subspaces that are inherently futile or fruitful for creativity, we need to incorporate semantic representations more directly. We therefore explore the use of distributional vector space models. Another goal of this section will be additional learning-based investigation to the compositional nature of creative word pairs, complementing the investigation in §4, which focused on the compositional aspect of creativity described in §2.1.

With above goals in mind, in what follows, we explore three different ways to learn compositional aspect of creative word pairs: (1) learning with explicit compositional vector operations (§5.1), (2) learning nonlinear composition via kernels (§5.2), (3) learning nonlinear composition via deep learning (§5.3). Note that in all these approaches, the notion of creative semantic subspace is integrated indirectly, as the feature representation always incorporates the resulting (composed) vector representations.

Baseline & Configuration We consider the concatenation of two word vectors $[\vec{w}_1; \vec{w}_2]$ as the baseline, since it can be viewed as what simple bag-of-word features would be. Since the size of creative pair dataset is not at scale yet, we choose to work with vector space models that are in reduced dimensions. We experimented with both Non-Negative Sparse Embedding (Murphy et al. (2012)) and neural semantic vectors of Huang et al. (2012), but report experiments with the latter only as those gave us slightly better results.

5.1 Compositional Vector Operations

We consider the following compositional vector operations inspired by recent studies for compositional distributional semantics (e.g., Guevara (2011), Clarke (2012), Mitchell and Lapata (2008), Widdows (2008)).

- **ADD**: $\vec{w}_1 + \vec{w}_2$
- **DIFF**: $abs(\vec{w}_1 - \vec{w}_2)$
- **MULT**: $\vec{w}_1 .* \vec{w}_2$
- **MIN**: $\min\{\vec{w}_1, \vec{w}_2\}$
- **MAX**: $\max\{\vec{w}_1, \vec{w}_2\}$

All operations take two input vectors $\in R^n$, and output a vector $\in R^n$. Each operation is applied element-wise. We then perform binary classification over the composed vectors using linear SVM. Besides using features based on the composed vectors, we also experiment with features based on concatenating multiple composed vectors, in the hope to capture more diverse compositional operations. See Table 5 for more details and experimental results.

5.2 Learning Nonlinear Composition via Kernels

As an alternative to explicit vector compositions, we also probe implicit operations based on non-linear combinations of semantic dimensions using kernels (e.g., Schölkopf and Smola (2002), Shawe-Taylor and Cristianini (2004)), in particular:

- Polynomial: $K(x, y) = (\gamma x^T y + r)^d$, $\gamma > 0$
- RBF: $K(x, y) = \exp(-\gamma \|x - y\|^2)$, $\gamma > 0$
- Laplacian: $K(x, y) = \exp(-\gamma \|x - y\|)$, $\gamma > 0$

5.3 Learning Non-linear Composition via Deep Learning

Yet another alternative to model non-linear composition is deep learning. To learn the non-linear transformation of a pair of semantic vectors, we explore the use of autoencoders (e.g., Pollack (1990), Voegtlin and Dominey (2005)). We follow the formulation of vector composition proposed by Socher et al. (2011) except that we do not stack autoencoders for recursion. More specifically, given the two input words $\vec{w}_1, \vec{w}_2 \in R^n$, we want to learn a vector space representation of their combination $\vec{p} \in R^n$. The recursive auto encoder (RAE) of Socher et al. (2011) models the composition of a word pair as a non-linear transformation of their concatenation $[\vec{w}_1; \vec{w}_2]$:

$$\vec{p} = f(M_1[\vec{w}_1; \vec{w}_2] + \vec{b}_1) \qquad (4)$$

where $M_1 \in R^{n \times 2n}$. After adding a bias term $\vec{b}_1 \in R^n$, a nonlinear element-wise function f such as $tanh$ is applied to the resulting vector. The representation $\vec{p}$ of the word pair is then fed into a reconstruction layer to reconstruct the two input vectors,

of Chen and Lin (2005) determines that the most two important ones are $RH(w_1, w_2)$ and $KL(w_1, w_2)$.

Methods	Accuracy
Creativity measures (§4.3)	
$\vec{F}_{12}$	62.30
Baseline: vector concatenation (no composition)	
$[\vec{w}_1; \vec{w}_2]$	<u>67.51</u>
Explicit vector composition (§5.1)	
$\vec{w}_1 + \vec{w}_2$	66.62
$abs(\vec{w}_1 - \vec{w}_2)$	60.03
$\min\{\vec{w}_1, \vec{w}_2\}$	66.08
$\max\{\vec{w}_1, \vec{w}_2\}$	64.97
$\vec{w}_1 .* \vec{w}_2$	56.34
$[abs(\vec{w}_1 - \vec{w}_2); \vec{w}_1; \vec{w}_2]$	**69.54**
$[\max\{\vec{w}_1, \vec{w}_2\}; \vec{w}_1; \vec{w}_2]$	68.02
Non-linear composition via kernels (§5.2)	
Polynomial	65.86
RBF	**69.16**
Laplacian	68.15
Non-linear composition via deep learning (§5.3)	
$f(M_1[\vec{w}_1; \vec{w}_2] + \vec{b}_1)$	67.25

Table 5: Performance comparison of creativity classifiers.

Incorrectly predicted word pairs	y^*	Semantically close word pairs	y^*
CONFUSION DUE TO WORD SIMILARITY (20/42)			
"entire carton"	-	*"whole angst"*	+
"outdated tax"	-	*"graconian tax"*	+
"dismissive way"	-	*"amorous way"*	+
"insidious part"	+	*"leather part"*	-
CONFUSION DUE TO SUBJECTIVE LABELING (8/42)			
"independent religion"	+	*"wonderful religion"*	-
WORD SENSE DISAMBIGUATION PROBLEMS (2/42)			
"fiscal cliff"	-	*"winding lake"*	+
"opera window"	+	*"work-shop floor"*	-

Table 6: Error analysis: y^* denotes the true label. For each incorrectly predicted word pair (left column), we show an example of semantically close word pairs (right column) with the opposite true label that might have confused learning.

and a softmax layer to predict the probability of the word pair being creative and not creative. We initialize the word vectors using the pre-learned vector space representations in Huang et al. (2012).

5.4 Experimental Results

Table 5 shows the performance comparison of different features sets and algorithms. In all cases, parameters are tuned from the training portion of the data. We see that simple vector composition alone does not perform better than vector concatenation $[\vec{w}_1; \vec{w}_2]$. However, combining $abs(\vec{w}_1 - \vec{w}_2)$ or $max\{\vec{w}_1; \vec{w}_2\}$ with $[\vec{w}_1; \vec{w}_2]$ perform better than concatenation. Kernels with non-linear transformation of feature space generally improve performance over linear SVM, suggesting that kernels capture some of the interesting compositional aspect of creativity that is not covered by some of the explicit vector compositions considered in §5.1. We also experimented with additional features driven from the creativity measures explored in §4, but we omit their results as those did not help improving the performance. Unfortunately learning nonlinear composition with deep learning did not yield better results. We conjecture that it is due to the small dataset we were able to obtain for this study, which may have not been enough to learn the rich parameter space of the nonlinear transformation matrix.

6 Analysis and Insight

Error analysis We manually inspected a randomly chosen 42 error cases, and characterize the potential causes of those errors. Examples of three types of errors are shown in Table 6. For each incorrectly predicted word pair, we also show a semantically close word pair with the opposite true label that might have confused the learning algorithm.

Visualization To gain additional insight, we project word pairs represented in their vector concatenations onto 2-dimensional space using t-Distributed Stochastic Neighbor Embedding (van der Maaten and Hinton (2008)). Figure 5 shows some of the interesting regions of the projection: some regions are relatively futile in having creative phrases (e.g., regions involving simple adjectives such as *"good"*, *"bad"*, regions corresponding to legal terms), while some regions are relatively more fruitful (e.g., regions involving abstract adjectives such as *"infinite"*, *"universal"*, *"fundamental"*). There are also many other regions (e.g., in the vicinity of *"true"*, *"perfect"* or *"intelligent"* in Figure 5) where the separation between creative and noncreative phrases are not as prominent. In those regions, compositional aspects would play a bigger role in determining creativity than memorizing fruitful semantic subspaces.

Figure 5: Creative (blue bold) and not creative (red italic) word pairs graph.

7 Related Work

Among computational approaches that touch on linguistic creativity, many focused on *metaphor* (e.g., Dunn (2013), Krishnakumaran and Zhu (2007), Mashal et al. (2007), Rumbell et al. (2008), Rentoumi et al. (2012), Mashal et al. (2009)). Other linguistic devices and phenomena related to creativity include *irony* (e.g., Davidov et al. (2010), González-Ibáñez et al. (2011), Filatova (2012)), *neologism* (e.g., Cartoni (2008)), *humor* (e.g., Mihalcea and Strapparava (2005), Purandare and Litman (2006)), and *similes* (e.g., Hao and Veale (2010)).

Veale (2011) proposed the new task of creative text retrieval to harvest expressions that potentially convey the same meaning as the query phrase in a fresh or unusual way. Our work contributes to the retrieval process of recognizing more creative phrases. Ozbal and Strapparava (2012) explored automatic creative naming of commercial products and services, focusing on the generation of creative phrases within a specific domain. Costello (2002) investigated the cognitive process that guides people's choice of words when making up a novel noun-noun compound. In contrast, we present a data-driven investigation to quantifying creativity in lexical composition. Memorability is loosely related to linguistic creativity (Danescu-Niculescu-Mizil et al. (2012)) as some of the creative quotes may be more memorable, but not all creative phrases are memorable and vice versa.

8 Conclusion

We presented the first study that focuses on learning and quantifying creativity in lexical compositions, exploring statistical techniques motivated by three different theories and hypotheses of creativity, ranging from divergent thinking, compositional structure, creative semantic subspace, and the connection to sentiment and connotation. Our experimental results suggest the viability of learning creative language, and point to promising directions for future research.

Acknowledgments This research was supported in part by the Stony Brook University Office of the Vice President for Research, and in part by gift from Google. We thank anonymous reviewers for insightful comments and suggestions.

References

T. Amabile. 1997. Motivating creativity in organizations: On doing what you love and loving what you do. *California Management Review*, 40(1):39–58.

Marco Baroni and Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1183–1193, Stroudsburg, PA, USA.

Thorsten Brants and Alex Franz. 2006. Web 1t 5-gram corpus version 1.1. *Google Inc*.

Ronald Carter and Michael McCarthy. 2004. Talking, creating: interactional language, creativity, and context. *Applied Linguistics*, 25(1):62–88.

Bruno Cartoni. 2008. Lexical resources for automatic translation of constructed neologisms: the case study of relational adjectives. In *LREC*.

Chih-Chung Chang and Chih-Jen Lin. 2011. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

Yi-Wei Chen and Chih-Jen Lin. 2005. Combining svms with various feature selection strategies. In *Taiwan University*. Springer-Verlag.

Noam Chomsky. 1965. *Aspects of the Theory of Syntax*, volume 11. The MIT press.

Carol Chomsky. 1976. Creativity and innovation in child language. *Journal of Education, Boston*.

Daoud Clarke. 2012. A context-theoretic framework for compositionality in distributional semantics. *Computational Linguistics*, 38(1):41–71.

W.M. Cohen and D.A. Levinthal. 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1).

Fintan J. Costello. 2002. Investigating creative language: People's choice of words in the production of novel noun-noun compounds. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*.

Arthur Cropley. 2006. In praise of convergent thinking. *Creativity Research Journal*, 18(3):391–404.

Cristian Danescu-Niculescu-Mizil, Justin Cheng, Jon Kleinberg, and Lillian Lee. 2012. You had me at hello: How phrasing affects memorability. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 892–901. Association for Computational Linguistics.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 107–116. Association for Computational Linguistics.

Jonathan Dunn. 2013. What metaphor identification systems can tell us about metaphor-in-language. *Meta4NLP 2013*, page 1.

Song Feng, Jun Sak Kang, Polina Kuznetsova, and Yejin Choi. 2013. Connotation lexicon: A dash of sentiment beneath the surface meaning. In *Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Sofia, Bulgaria, Angust. Association for Computational Linguistics.

Elena Filatova. 2012. Irony and sarcasm: Corpus generation and analysis using crowdsourcing. In *LREC*, pages 392–398.

Stefan L Frank. 2010. Uncertainty reduction as a measure of cognitive processing effort. In *Proceedings of the 2010 workshop on cognitive modeling and computational linguistics*, pages 81–89. Association for Computational Linguistics.

Sigmund Freud. 1908. Creative writers and daydreaming. *Standard edition*, 9:143–153.

Susan R. Fussell and Mallie M. Moss. 1998. Figurative language in descriptions of emotional states. In *Social and cognitive approaches to interpersonal communication*.

Pablo Gervás. 2010. Engineering linguistic creativity: Bird flight and jet planes. In *Proceedings of the NAACL HLT 2010 Second Workshop on Computational Approaches to Linguistic Creativity*, pages 23–30. Association for Computational Linguistics.

Katie Glaskin. 2011. Dreams, memory, and the ancestors: creativity, culture, and the science of sleep. *Journal of the royal anthropological institute*, 17(1):44–62.

Andrew Goatly. 1997. *The language of metaphors*. Routledge.

Roberto González-Ibáñez, Smaranda Muresan, and Nina Wacholder. 2011. Identifying sarcasm in twitter: A closer look. In *ACL (Short Papers)*, pages 581–586. Citeseer.

Emiliano Guevara. 2011. Computing semantic compositionality in distributional semantics. In *Proceedings of the Ninth International Conference on Computational Semantics (IWCS 2011)*, pages 135–144. Citeseer.

John Hale. 2003. The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32(2):101–123.

John Hale. 2006. Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4):643–672.

Yanfen Hao and Tony Veale. 2010. An ironic fist in a velvet glove: Creative mis-representation in the construction of ironic similes. *Minds and Machines*, 20(4):635–650.

Eric H. Huang, Richard Socher, Christopher D. Manning, and Andrew Y. Ng. 2012. Improving Word Representations via Global Context and Multiple Word Prototypes. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Saisuresh Krishnakumaran and Xiaojin Zhu. 2007. Hunting elusive metaphors using lexical resources. In *Proceedings of the Workshop on Computational approaches to Figurative Language*, pages 13–20. Association for Computational Linguistics.

George Lakoff and Mark Johnson. 1980. *Metaphors we Live by*. University of Chicago Press, Chicago.

N. Mashal, M. Faust, T Hendler, and M. Jung-Beeman. 2007. An fmri investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain and Language*, pages 115 – 126.

N Mashal, M Faust, T Hendler, and M Jung-Beeman. 2009. An fmri study of processing novel metaphoric sentences. *Laterality*, (1):30–54.

Janet Maybin and Joan Swann. 2007. Everyday creativity in language: Textuality, contextuality, and critique. *Applied Linguistics*, 28(4):497–517.

Robert R McCrae. 1987. Creativity, divergent thinking, and openness to experience. *Journal of personality and social psychology*, 52(6):1258.

Susan M. McCurry and Steven C. Hayes. 1992. Clinical and experimental perspectives on metaphorical talk. *Clinical Psychology Review*, 12(7):763 – 785.

Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 531–538, Vancouver, British Columbia, Canada, October. Association for Computational Linguistics.

Jeff Mitchell and Mirella Lapata. 2008. Vector-based models of semantic composition. In *In Proceedings of ACL-08: HLT*, pages 236–244.

Jeff Mitchell and Mirella Lapata. 2009. Language models based on semantic composition. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 430–439. Association for Computational Linguistics.

Brian Murphy, Partha Pratim Talukdar, and Tom Mitchell. 2012. Learning effective and interpretable semantic models using non-negative sparse embedding. In *COLING*, pages 1933–1950.

Edward Necka. 1999. Memory and creativity. *Encyclopedia of creativity, ed. by MA Runco, SR Pritzker*, 2:193–99.

Gozde Ozbal and Carlo Strapparava. 2012. A computational approach to the automation of creative naming. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 703–711, Jeju Island, Korea, July. Association for Computational Linguistics.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics.

J. B. Pollack. 1990. Recursive distributed representation. *Artificial Intelligence*, 46:77–105.

Amruta Purandare and Diane Litman. 2006. Humor: Prosody analysis and automatic recognition for f* r* i* e* n* d* s*. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 208–215. Association for Computational Linguistics.

Vassiliki Rentoumi, George A. Vouros, Vangelis Karkaletsis, and Amalia Moser. 2012. Investigating metaphorical language in sentiment analysis: A sense-to-sentiment perspective. *ACM Trans. Speech Lang. Process.*, 9(3):6:1–6:31, November.

Sebastian Rudolph and Eugenie Giesbrecht. 2010. Compositional matrix-space models of language. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 907–916. Association for Computational Linguistics.

Tim Rumbell, John Barnden, Mark Lee, and Alan Wallington. 2008. Affect in metaphor: Developments with wordnet.

Bernhard Schölkopf and Alexander J Smola. 2002. *Learning with kernels*. The MIT Press.

John Shawe-Taylor and Nello Cristianini. 2004. *Kernel methods for pattern analysis*. Cambridge university press.

Ekaterina Shutova. 2010. Models of metaphor in nlp. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 688–697, Stroudsburg, PA, USA. Association for Computational Linguistics.

Rion Snow, Brendan O'Connor, Daniel Jurafsky, and Andrew Y Ng. 2008. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the conference on empirical methods in natural language processing*, pages 254–263. Association for Computational Linguistics.

Richard Socher, Jeffrey Pennington, Eric H Huang, Andrew Y Ng, and Christopher D Manning. 2011. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 151–161. Association for Computational Linguistics.

L.J.P. van der Maaten and G.E. Hinton. 2008. Visualizing high-dimensional data using t-sne.

Tony Veale, Yanfen Hao, and Guofu Li. 2008. Multilingual harvesting of cross-cultural stereotypes. In *ACL*, pages 523–531.

Tony Veale. 2011. Creative language retrieval: A robust hybrid of information retrieval and linguistic creativity. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 278–287, Portland, Oregon, USA, June. Association for Computational Linguistics.

Thomas Voegtlin and Peter F. Dominey. 2005. Linear recursive distributed representations. *Neural Netw.*, 18(7):878–895, September.

Dominic Widdows. 2008. Semantic vector products: Some initial investigations. In *Proceedings of the Second AAAI Symposium on Quantum Interaction*.

Theresa Wilson, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. Opinionfinder: A system for subjectivity analysis. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, pages 34–35. Association for Computational Linguistics.

Ainur Yessenalina and Claire Cardie. 2011. Compositional matrix-space models for sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 172–182. Association for Computational Linguistics.

Xiaojin Zhu, Zhiting Xu, and Tushar Khot. 2009. How creative is your writing? a linguistic creativity measure from computer science and cognitive psychology perspectives. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*, pages 87–93. Association for Computational Linguistics.