Idiom Paraphrases: Seventh Heaven vs Cloud Nine

Maria Pershina **Yifan He** **Ralph Grishman**
Computer Science Department
New York University
New York, NY 10003, USA
{pershina,yhe,grishman}@cs.nyu.edu

Abstract

The goal of paraphrase identification is to decide whether two given text fragments have the same meaning. Of particular interest in this area is the identification of paraphrases among short texts, such as SMS and Twitter. In this paper, we present idiomatic expressions as a new domain for short-text paraphrase identification. We propose a technique, utilizing idiom definitions and continuous space word representations that performs competitively on a dataset of 1.4K annotated idiom paraphrase pairs, which we make publicly available for the research community.

1 Introduction

The task of paraphrase identification, i.e. finding alternative linguistic expressions of the same or similar meaning, attracted a great deal of attention in the research community in recent years (Bannard and Callison-Burch, 2005; Sekine, 2005; Socher et al., 2011; Guo et al., 2013; Xu et al., 2013; Wang et al., 2013; Zhang and Weld, 2013; Xu et al., 2015).

This task was extensively studied in Twitter data, where millions of user-generated tweets talk about the same topics and thus present a natural challenge to resolve redundancy in tweets for many applications, such as textual entailment (Zhao et al., 2014), text summarization (Lloret et al., 2008), first story detection (Petrovich, 2012), search (Zanzotto et al., 2011), question answering (Celikyilmaz, 2010), etc.

In this paper we explore a new domain for the task of paraphrase identification - idiomatic expressions, in which the goal is to determine whether two idioms convey the same idea.

This task is related to previous short-text paraphrase tasks, but it does not have access to many of the information sources that can be exploited in Twitter/short text paraphrasing: unlike tweets, idioms do not have hashtags, which are very strong topic indictors; unlike SMS, idioms do not have timestamp or geographical metadata; and unlike news headlines, there are no real world events that can serve as anchors to cluster similar expressions. In addition, an idea, or a moral of the idiom is often expressed in an indirect way, e.g. the idioms

(1) make a mountain out of a molehill
(2) tempest in a teapot

convey similar ideas[1]:

(1) If somebody makes a mountain out of a molehill they exaggerate the importance or seriousness of a problem.
(2) If people exaggerate the seriousness of a situation or problem they are making a tempest in a teapot.

There is a line of research focused on extracting idioms from the text or identifying whether a particular expression is idiomatic (or a non-compositional multi-word expression) (Muzny and Zettlemoyer, 2013; Shutova et al., 2010; Li and Sporleder, 2009; Gedigian et al., 2006; Katz and Giesbrecht, 2006). Without linguistic sources such as Wiktionary, usingenglish.com, etc, it is often hard to understand what the meaning of a particular idiom is. It is even harder to determine whether two idioms convey the same idea or find alternative idiomatic expressions. Using idiom definitions, given by linguistic resources, one can view this problem as identifying paraphrases between definitions and thus deciding on paraphrases between corresponding idioms. Efficient techniques for identifying idiom paraphrases would complement any paraphrase identification system, and thus improve the downstream applications, such as question answering, summariza-

[1]Definitions of these idioms are taken from http://www.usingenglish.com

Proceedings of the EMNLP 2015 Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics, pages 76–82,
Lisboa, Portugal, 18 September 2015.

tion, opinion mining, information extraction, and machine translation.

To the best of our knowledge we are the first to address the problem of determining whether two idioms convey the same idea, and to propose a new scheme that utilizes idiom definitions and continuous space word representation (word embedding) to solve it. By linking word- and sentence-level semantics our technique outperforms state-of-the-art paraphrasing approaches on a dataset of 1.4K annotated idiom pairs that we make publicly available.

2 Related Work

There is no strict definition of a paraphrase (Bhagat and Hovy, 2013) and in linguistic literature paraphrases are most often characterized by an approximate equivalence of meanings across sentences or phrases.

A growing body of research investigates ways of paraphrase detection in both supervised (Qiu et al., 2006; Wan et al., 2006; Das and Smith, 2009; Socher et al., 2011; Blacoe and Lapata, 2012; Madnani and Tetreault, 2012; Ji and Eisenstein, 2013) and unsupervised settings (Bannard and Callison-Burch, 2005; Mihalcea et al., 2006; Rus et al., 2008; Fernando and Stevenson, 2008; Islam and Inkpen, 2007; Hassan and Mihalcea, 2011). These methods mainly work on large scale news data. News data is very different from ours in two aspects: most news text can be interpreted literally and similar news events (passing a legislation, death of a person, elections) happen repeatedly. Therefore, lexical anchors or event anchors can work well on news text, but not necessarily on our task.

Millions of tweets generated by Twitter users every day provide plenty of paraphrase data for NLP research. An increasing interest in this problem led to the **P**araphrase and Semantic Similarity **I**n **T**witter (PIT) task in SemEval-2015 competition (Xu et al., 2015). Existing bias towards Twitter paraphrases results in sophisticated systems that exploit character level similarity or metadata. But models relying on these insights are not necessarily applicable to other domains where misspellings are rare, or metadata is not available.

Idiomatic expressions constitute an essential part of modern English. They often behave idiosyncratically and are therefore a significant challenge for natural language processing systems. Recognizing when two idiomatic expressions convey similar ideas is crucial to recognizing the sentiment of the author, identifying correct triggers for events, and to translating the idiom properly. However, although there are several existing models to identify paraphrases in short text, idioms have very different characteristics from the data that those models are built on. In this paper, we experiment with two state-of-the-art paraphrasing models that are outperformed on our dataset of idiomatic expressions by a simple technique, raising a question on how well existing paraphrase models generalize to new data.

3 The Challenge

Identifying idiom paraphrases is an interesting and challenging problem. Lexical similarity is not a reliable clue to find similar idioms. Some idioms look very similar, differ in only one or two words, and convey the same idea. For example, "like two peas in a pod" vs "like peas in a pod" ("if people or things are like peas in a pod they look identical"), but other idioms that look similar can have very different meaning, e.g. "well oiled" vs "well oiled machine" ("if someone is well oiled they have drunk a lot" vs "something that functions very well is a well oiled machine").

Finally, there are idioms that do not have any words in common at all and may seem quite different for a person not familiar with idiomatic expressions, but still have similar meaning. For example, "cross swords" vs "lock horns" ("when people cross swords they argue or dispute" vs "when people lock horns they argue or fight about something"). Thus, a natural way to identify idiom paraphrases is to focus on idiom definitions that explain meaning of an idiom in a clear and concise way.

4 Lexical vs Semantic Similarities

Our dataset consists of pairs $\langle idiom, definition\rangle$.

We use two types of similarity measures to compute how similar definitions of different idioms are: the $lexical$ similarity is based on a lexical (word) overlap between two definitions, and the $semantic$ similarity captures the overall semantic meaning of the whole sentence.

Lexical similarity. We compute cosine similarity between vectors $\overrightarrow{v}_{d_1}$ and $\overrightarrow{v}_{d_2}$, representing idiom descriptions d_1 and d_2 and weight each word in these vectors by its tf-idf score:

$$lexSim(d_1, d_2) = cosine(\overrightarrow{v}_{d_1}, \overrightarrow{v}_{d_2}), \quad (1)$$

where $\overrightarrow{v}_d$ is a $|V|$-dimensional vector with V being the vocabulary of all definition words.

Semantic similarity. To capture the overall meaning of the definitions d we combine word embeddings (Collobert et al., 2011; Turian et al., 2010) for all words in d using two combination schemes:

- Averaged sum:

$$\overrightarrow{averaged_d} = \frac{1}{|d|} \sum_{word \in d} \overrightarrow{emb}(word) \quad (2)$$

- Weighted sum:

$$\overrightarrow{weighted_d} = \quad (3)$$

$$\frac{1}{\sum_{word \in d} \text{tfidf}_{word}} \sum_{word \in d} \text{tfidf}_{word} \cdot \overrightarrow{emb}(word)$$

Then semantic similarity is measured as

$$semSim(d_1, d_2) = cosine(comb_{d_1}, comb_{d_2}) \quad (4)$$

where $\overrightarrow{comb_d}$ is a 100-dimensional vector combined from word embeddings $\overrightarrow{emb}(word)$ (Turian et al., 2010) for words in description d using either averaged (2) or weighted (3) combination schemes. [2]

4.1 IdiomSim

There is a tradeoff between the two similarity measures $lexSim$ and $semSim$ (Section 4): while the first one captures the actual lexical overlap, the second one can better capture the closeness in semantic meaning. To find an optimal balance between the two we consider their weighted sum

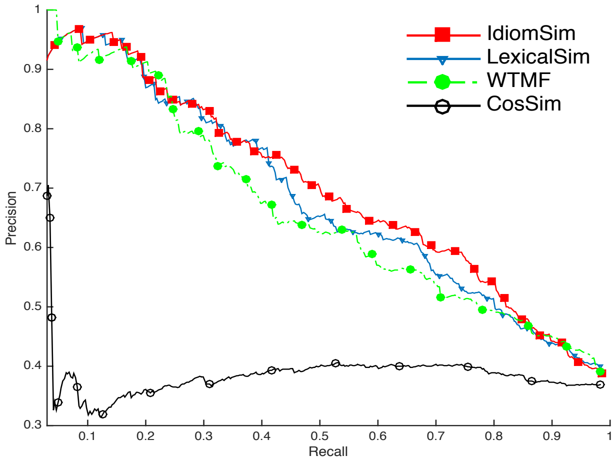
$$IdiomSim(d_1, d_2) = \quad (5)$$

$$(1 - \alpha) \cdot lexSim(d_1, d_2) + \alpha \cdot semSim(d_1, d_2)$$

and decide on an α by optimizing for a maximal F-score on a development dataset.

5 Experiments

Data. We collected 2,432 idioms from http://www.usingenglish.com, a site for English learners, where every idiom has a unique description giving a clear explanation of the idiom's meaning. As opposed to tweets there are no hashtags, no topics or trends, no timestamps, or any other default evidence, that two idioms may convey similar ideas. Thus it becomes a challenging



od	1		
o Sim		1	
S	1		1
	1	1	
i Sim			
IdiomSim^{ave}			
IdiomSim			
IdiomSim+			1

Figure 1: Comparison of IdiomSim with baselines CosSim, LexicalSim, and state-of-the-art paraphrasing models: ASOBEK, WTMF.

task itself to construct a dataset of pairs that is guaranteed to have a certain fraction of true paraphrases.

We used a simple cosine similarity between all possible idiom definitions pairs to have a ranked list and labeled the top 1.5K pairs. Three annotators were asked to label each pair of idiom definitions as "similar" (score 2), "have something in common" (score 1), "not similar" (score 0). 0.1K pairs received a total score of 4 (either 2+2+0, or 2+1+1), and were further removed as debatable. The rest of the labeled pairs were randomly split into 1K for test data and 0.4K for development. Only pairs that received a total score of 5 or higher were considered as positive examples. There are 364 and 96 true paraphrases in our test and development sets respectively. [3]

Baselines. Our baselines are simple and tf-idf weighted cosine similarity between idiom description sentences: CosSim and LexicalSim.

We compare our method with the deterministic state-of-the-art ASOBEK model (Eyecioglu and

[2] We use 100-dimensional Turian word embeddings available at `http://metaoptimize.com/projects/wordreprs/`

[3] `https://github.com/masha-p/Idiom_Paraphrases`

Idioms	Descriptions
seventh heaven cloud nine	if you are in seventh heaven you are extremely happy if you are on cloud nine you are extremely happy
face only a mother could love stop a clock	when someone has a face only a mother could love they are ugly a face that could stop a clock is very ugly indeed
take your medicine face the music	if you take your medicine you accept the consequences of something you have done wrong if you have to face the music you have to accept the negative consequences of something you have done wrong
well oiled drunk as a lord	if someone is well oiled they have drunk a lot someone who is very drunk is as drunk as a lord
cheap as chips to be dog cheap	if something is very inexpensive it is as cheap as chips if something is dog cheap it is very cheap indeed
great minds think alike on the same wavelength	if two people have the same thought at the same time if people are on the same wavelength they have the same ideas and opinions about something
could eat a horse hungry as a bear	if you are very hungry you could eat a horse if you are hungry as a bear it means that you are really hungry
cross swords lock horns	when people cross swords they argue or dispute when people lock horns they argue or fight about something
talk the hind legs off a donkey talk the legs off an iron pot	a person who is excessively or extremely talkative can talk the hind legs off a donkey somebody who is excessively talkative or is especially convincing is said to talk the legs off an iron pot

Table 1: Examples of extracted idiom paraphrases.

Keller, 2015) that was ranked first among 19 teams in the Paraphrase in Twitter (PIT) track on the SemEval 2015 shared task (Xu et al., 2015). This model extracts eight simple and elegant character and word features from two sentences to train an SVM with linear kernel. It achieves an F-score of 55.1% on our test set.[4]

We also compare our method with the state-of-the-art Weighted Textual Matrix Factorization model (WTMF) (Guo et al., 2013),[5] which is specifically developed for short sentences by modeling the semantic space of words, that can be either present or absent from the sentences (Guo and Diab, 2012). This model achieves a maximal F-score of 61.4% on the test set.

The state-of-the-art model for lexically divergent paraphrases on Twitter (Xu et al., 2015) is tailored for tweets and requires topic and anchor words to be present in the sentence, which is not applicable to idiom definitions.

Evaluation and Results. To evaluate models we plot precision-recall curves for CosSim, WTMF, LexicalSim, and IdiomSim (for clarity we omit curves for other models). We also compare maximal F-score for all models. We observe that simple cosine similarity (CosSim) achieves a maximal F-score of 53.7%, LexicalSim is a high baseline and achieves an F-score of 63.75%. When we add averaged word embeddings the maximal F-score is 64.4% (IdiomSim$^{\text{ave}}$). With tfidf weighted word embeddings we achieve F-score of 65.9% (IdiomSim). By filtering out uninformative words such as “a”, “the”, etc (12 words total) we improve the F-score to 66.6% (IdiomSim+), outperforming state-of-the-art paraphrase models by more than 5% absolute (Figure 1). Both IdiomSim and IdiomSim+ outperform WTMF significantly according to a paired t-test with p less than 0.05.

Examples and Discussion. We use threshold, corresponding to a maximal F-score obtained on the development dataset, and explore paraphrases from test dataset scored higher and lower than this threshold. Examples of extracted idiom paraphrases are in Table 1. Examples of false positives and false negatives are in Table 2.

Simple word overlap is not a reliable clue to de-

[4]We thank Asli Eyecioglu for running her ASOBEK model on our test data.

[5]The source code for WTMF is available at `http://www.cs.columbia.edu/~weiwei/code`

Idioms	Descriptions
False positives	
healthy as a horse	if you are as healthy as a horse you are very healthy
an apple a day keeps the doctor away	eating healthy food keeps you healthy
jersey justice	jersey justice is a very severe justice
justice is blind	justice is blind means that justice is impartial and objective
heart of steel	when someone has a heart of steel they do not show emotion or are not affected emotionally
heart of glass	when someone has a heart of glass they are easily affected emotionally
False negatives	
like a kid in a candy store	if someone is like a kid in a candy store they are very excited about something
bee in your bonnet	if someone is very excited about something they have a bee in their bonnet
easy as falling off a log	something very easy or simple to do is as easy as falling off a log
no sweat	no sweat means something is easy
hopping mad	if you are hopping mad you are extremely angry
off on one	if someone goes off on one they get extremely angry indeed

Table 2: Examples of false positive and false negative paraphrases.

cide on a paraphrase between two idiom descriptions. Since words are main units in the computation (5) our metric is biased towards lexical similarity. Thus we get a false positive paraphrase between "healthy as a horse" and "an apple a day". The first one is rather a statement about someone's health while the second one is an advice on how to be healthy. Moreover, idioms "heart of steel" vs "heart of glass" convey opposite ideas of being "not affected emotionally" vs being "easily affected emotionally". Having "heart" and "affected emotionally" in both idiom descriptions leads to a high cosine similarity between them and results in a false positive decision. For the same reason lexically divergent idiom descriptions get a lower rank while convey similar ideas, e.g. "hopping mad" vs "off on one".

Combining lexical and sentence similarity via (5) performs better than lexical similarity alone (Figure 1) but still does not capture all aspects of a true paraphrase.

6 Conclusion and Future Work

In this paper we present a new domain for the paraphrase identification task: to find paraphrases among idiomatic expressions. We propose a simple scheme to compute the similarity of two idiom definitions that outperforms state-of-the-art paraphrasing models on the dataset of idiom paraphrases that we make publicly available.

Our future work will be focused on exploring different strategies to compute semantic similarity between sentences, developing a comprehensive idiom similarity measure that will utilize both idioms and their definitions, and on comparing text with an idiom and a general text as a realistic scenario for paraphrase identification. It is a new and a challenging task and thus opens up many opportunities for further research in paraphrase identification and all its downstream applications.

Acknowledgments

We thank Thien Huu Nguyen of New York University and Asli Eyecioglu of University of Sussex for their help and advice.

References

Eneko Agirre, Mona Diab, Daniel Cer, and Aitor Gonzalez-Agirre. (2012). Semeval-2012 task 6: A pilot on semantic textual similarity. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*SEM).*

Eneko Agirre, Carmen Banea, Claire Cardie, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, Weiwei Guo, Inigo Lopez-Gazpio, Montse Maritxalar, Rada Mihalcea, German Rigau, Larraitz Uria, and Janyce Wiebe. (2015). Semeval-2015 task 2: Semantic textual similarity, English, Spanish and Pilot on Interpretability. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval).*

Colin Bannard and Chris Callison-Burch. (2005). Paraphrasing with Bilingual Parallel Corpora. In *Proceedings of the 43th Annual Meeting of the Association for Computational Linguistics (ACL).*

Marco Baroni, Georgiana Dinu, and German Kruszewski. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics (ACL).*

Rahul Bhagat and Eduard Hovy. (2013). What is a paraphrase? In *Proceedings of the International Conference on Computational Linguistics (COLING).*

Julia Birke and Anoop Sarkar. (2006). A clustering approach for nearly unsupervised recognition of nonliteral language. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL).*

William Blacoe and Mirella Lapata. (2012). A comparison of vector-based representations for semantic composition. In *Proceedings of EMNLP-CoLNN.*

Asli Celikyilmaz, Dilek Hakkani-Tur, and Gokhan Tur. (2010). LDA based similarity modeing for question answering. In *Proceedings of the NAACL HLT 2010 Workshop on Semantic Search.*

Ronan Collobert, Jason Weston, Leon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa (2011). Natural Language Processing (Almost) from Scratch. In *Journal of Machine Learning Research (JMLR).*

Dipanjan Das and Noah A. Smith. (2009). Paraphrase identification as probabilistic quasi-synchronous recognition. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language (ACL-IJCNLP).*

Asli Eyecioglu and Bill Keller. (2015). ASOBEK: Twitter Paraphrase Identification with Simple Overlap Features and SVMs In *Proceedings of 9th International Workshop on Semantic Evaluation (SemEval).*

Samuel Fernando and Mark Stevenson (2008). A semantic similarity approach to paraphrase detection. *Computational Linguistics UK (CLUK) 11th Annual Research Colloquium.*

Matt Gedigian, John Bryant, Srini Narayanan, and Branimir Ciric. (2006). Catching metaphors. In *Proceedings of the Third Workshop on Scalable Natural Language Understanding (ScaNaLU).*

Weiwei Guo and Mona Diab. (2013). Modeling Sentences in the Latent Space. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL).*

Weiwei Guo, Hao Li, Heng Ji, and Mona Diab. (2013). Linking Tweets to News: A Framework to Enrich Short Text Data in Social Media. In *Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics (ACL).*

Samer Hassan and Rada Mihalcea. (2011). Semantic relatedness using salient semantic analysis. In *Proceedings of the twenty-fifth Association for the Advancement of Artificial Intelligence Conference (AAAI).*

Aminul Islam and Diana Inkpen. (2007). Semantic similarity of short texts. In *Proceedings of Conference on Recent Advances in Natural Language Processing (RANLP).*

Yangfeng Ji and Jacob Eisenstein. (2013). Discriminative improvements to distributional sentence similarity. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).*

Graham Katz and Eugenie Giesbrecht. (2006). Automatic identification of non-compositional multiword expressions using latent semantic analysis. In *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties (MWE).*

Linlin Li and Caroline Sporleder. (2009). Classifier combination for contextual idiom detection without labeled data. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).*

Elena Lloret, Oscar Ferrandez, Rafael Munoz, and Manuel Palomar. (2008). A text summarization approach under the influence of textual entailment. In *Proceedings of the 5th International Workshop on Natural Language Processing and Cognitive Science (NLPCS).*

Nitin Madnani and Joel Tetreault. (2012). Reexamining machine translation metrics for paraphrase identification. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT).*

Rada Mihalcea, Courtney Corley, and Strapparava. (2006). Corpus-based and knowledge-based measures of text semantic similarity. In *Proceedings of the Association for the Advancement of Artificial Intelligence Conference (AAAI).*

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. (2013). Efficient estimation of word representations in vector space. In *Proceedings of Workshop at the International Conference on Learning Representations (ICLR).*

Grace Muzny and Luke Zettlemoyer. (2013). Automatic Idiom Identification in Wiktionary. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP).*

Sasa Petrovic, Miles Osborne, and Victor Lavrenko (2012). Using paraphrases for improving first story detection in news and Twitter. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT)*.

Long Qiu, Min-Yen Kan, and Tat-Seng Chua. (2006). Paraphrase recognition via dissimilarity significance classification. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP)*.

Vasile Rus, Philip M. McCarthy, Mihai C. Lintean, Danielle S. McNamara, and Arthur C. Graesser (2008). Paraphrase identification with lexico-syntactic graph subsumption. In *Proceedings of the Twenty-First International FLAIRS Conference*.

Satoshi Sekine, (2005). Automatic paraphrase discovery based on context and keywords between NE pairs. In *Proceedings of the 3rd International Workshop on Paraphrasing*.

Yusuke Shinyama, Satoshi Sekine, and Kiyoshi Sudo. (2002). Automatic paraphrase acquisition from news articles. In *Proceedings of the 2nd International Conference on Human Language Technology Research (HLT)*.

Ekaterina Shutova, Lin Sun, and Anna Korhonen. (2010). Metaphor identification using verb and noun clustering. In *Proceedings of the International Conference on Computational Linguistics (COLING)*.

Richard Socher, Eric H Huang, Jeffrey Pennington, Andrew Y Ng, and Christopher D Manning. (2011). Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*.

Joseph Turian, Lev Ratinov, and Yoshua Bengio. (2010). Word representations: A simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Stephen Wan, Mark Dras, Robert Dale, and Cecile Paris. (2006). Using dependency-based features to take the parafarce out of paraphrase. In *Proceedings of the Australasian Language Technology Workshop*.

Ling Wang, Chris Dyer, Alan W Black, and Isabel Trancoso. (2013). Paraphrasing 4 microblog normalization. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP)*.

Wei Xu, Chris Callison-Burch, and William B. Dolan. (2015). SemEval-2015 Task 1: Paraphrase and Semantic Similarity in Twitter (PIT). In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval)*.

Wei Xu, Alan Ritter, Chris Callison-Burch, William B. Dolan, and Yangfeng Ji. (2015). Extracting Lexically Divergent Paraphrases from Twitter. *Transactions of the Association for Computational Linguistics (TACL)*.

Wei Xu, Alan Ritter, and Ralph Grishman. (2013). Gathering and Generating Paraphrases from Twitter with Application to Normalization. In *Proceedings of the Sixth Workshop on Building and Using Comparable Corpora (BUCC)*.

Fabio Massimo Zanzotto, Marco Pennacchiotti, and Kostas Tsioutsiouliklis (2011). Linguistic redundancy in Twitter. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP)*.

Congle Zhang and Daniel S. Weld (2013). Harvesting parallel news streams to generate paraphrases of event relations. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP)*.

Jiang Zhao, Man Lan, Zheng-Yu Niu, and Dong-Hong Ji. (2014). Recognizing cross-lingual textual entailment with co-training using similarity and difference views. In *Proceedings of International Joint Conference on Neural Networks (IJCNN)*.

Jiang Zhao, Man Lan, and Jun Feng Tian. (2015). ECNU: Using Traditional Similarity Measurements and Word Embedding for Semantic Textual Similarity Estimation. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval)*.