A Multi-Dictionary Approach to Abstractness/Concreteness-Based Authorship Attribution

Lubomir Ivanov
Iona University, New Rochelle, NY 10801
livanov@iona.edu

Abstract

We present some early results from a research project aimed at exploring the usefulness of abstractness/concreteness as stylistic features for authorship attribution. We conjecture that authors use abstract/concrete words and phrases in sufficiently unique ways, so that machine learning classifiers can learn to distinguish the individual authors' writing styles. Our approach is based on using the abstractness ratings of words and phrases from texts with established authorship to generate training vectors for different machine learning classifiers. The combined word/phrase ratings are extracted from two separate abstractness dictionaries – an approach that yields stronger results than using single abstractness dictionaries. The paper describes the details of our methodology and compares the results to those obtained using traditional authorship attribution stylistic features. The limitations of our current methodology and directions for further research are outlined at the end of the paper.

Introduction

The goal of authorship attribution is to determine, with a high degree of confidence, the identity of the author of a text or a document whose authorship is either unknown or disputed. Attribution is based on the use of a machine learning model or an ensemble of models, trained to recognize a set of stylistic features. These features are used in a unique way by each author and, collectively, define the author's writing style. The accuracy of attribution depends on a number of factors such as the choice of machine learning algorithms, the availability of a sufficiently large corpus of verified works by the candidate authors, and the selection of an appropriate set of stylistic features. Among the most commonly used features, which have consistently yielded strong results are function words, character- and word-n-grams, part-of-speech (PoS) tags, prepositions, suffixes, etc. For an overview of authorship attribution, the reader is referred to (Juola 2009, Stamatatos 2009 & 2016).

In recent years, it has been demonstrated that an ensemble of classifiers, using a broad selection of stylistic features, significantly outperforms individual classifier/feature pairs (Petrovic I. 2019, Petrovic S. 2018). This has motivated research into new, non-traditional stylistic features, which complement the primary features listed above and strengthen the overall accuracy of the attribution hypothesis. Recent examples of such novel stylistic features include prosodic features such as lexical stress, assonance, consonance, and alliteration. They have been shown to raise the accuracy of the attribution task when used in combination with traditional stylistic features (Ivanov 2019; Ivanov, Aebig, and Meerman 2018, Ivanov 2016, Ivanov and Petrovic 2015). Our work seeks to enhance this set of alternative stylistic features by exploring the usefulness of abstractness/concreteness of words and phrases as stylistic features for authorship attribution. The initial motivation for this work came from discussions with humanities colleagues, who suggested that certain authors tend to use abstract terms and ideas much more often than other writers. The primary goal of our research is to determine if such differences in the use of abstractness/concreteness, once quantified, are sufficiently large and unique to allow the automatic identification of an author's writing style.

This paper describes our current methodology and presents some early, promising results. We begin by outlining the psycholinguistic foundations of the concepts of abstractness and concreteness and briefly consider previous attempts to capture their essence computationally. We then describe our abstractness/concreteness-based authorship attribution approach and present some results from experiments performed using the Reuters-RCV1 dataset. The results are compared to those obtained using traditional stylistic features. Finally, we discuss some current limitations of our methodology and outline directions for further research.

Abstractness/Concreteness

The concepts of abstractness and concreteness are intuitively understood by most people, but difficult to define formally. Abstractness and concreteness have been studied from a philosophical, psychological, neuro-physiological, linguistic, and literary perspectives. Given our focus on authorship attribution, we limit our attention to the abstractness/concreteness of words and phrases in text.

In psycholinguistic literature, the most common way to define the concreteness of a word/phrase is based on our sensory ability to perceive or experience the object or phenomenon described by that word/phrase. Abstractness is the opposite of concreteness. A different way to define abstractness is in terms of "generality": Abstract terms tend to be more broadly defined, more general. According to (Burgoon, Henderson, and Markman 2013), abstractness is "a process of information reduction that allows for efficient storage and retrieval of central knowledge".

Much work has been carried out on the computational aspects of concreteness/abstractness recognition and classification, including metaphor-, hyperbole- and idiom detection and rating (Spreen & Schulz 1966, Paivio et al 1968, Coltheart 1981, Birke & Sarkar 2006 & 2007, Sporleder & Li 2009, Turney et al 2011, Kwong 2011, Assaf et al 2013, Neuman et al 2013, Shutova et al 2013, Brysbaert et al 2014, Tsvetkov et al 2014, Klebanov et al 2015, Rei et al 2017, Wu et al 2018, Gao et al 2018, Reif et al 2019).

An early attempt to capture abstractness/concreteness was presented by Spreen and Schulz in their 1966 seminal paper. The authors provided 7-point concreteness ratings for 329 nouns. They also compared abstractness defined in terms of (lack of) sensory perception with "generality" abstractness, finding a strong correlation between the two. Later, ratings collected by Paivio, Yuille, and Madigan were added to the Spreen/Schulz data and made available in the 4292-word MRC database (Coltheart, 1981), which also provided "imageability" ratings for 8900 words.

In 2011, Turney, Neuman, Assaf, and Cohen presented a classification of English verbs and adjective-noun combinations as literal or metaphorical based on the assumption that the metaphorical word use is related to its abstractness.

In 2014, Tsvetkov, Boytsov, Gershman, Nyberg, and Dyer presented a language-independent metaphor-detecting classifier, trained on English words. The training was based on WordNet categories and vector-space word representations. Of course, WordNet, itself, is based on the hypo-/hypernymy relation, which captures the essence of concreteness/abstractness in terms of "generalization".

In 2014, Brysbaert, Warriner, and Kuperman introduced a collection of 37058 words and 2896 two-word phrases rated for concreteness by 4000 evaluators (approximately 25 evaluators per word). The rating uses a 5-point scale, where lower values indicate abstractness and higher values - concreteness. For each word/phrase, the mean rating and the standard deviation (i.e., evaluator agreement) were recorded. The BWK (a.k.a. "Brysbaert") dictionary has become a de-facto standard for research in the application of abstractness/concreteness. Among the recent uses of the BWK dictionary was the training of machine learning models to automatically rate word/phrase abstractness (Köper and Schulte im Walde 2016, Maudslay et al 2020).

In 2018, Rabinovich et al introduced an abstractness dictionary of 300K words and phrases. The dictionary contains single words as well as two- and three word-grams. The abstractness rating of each dictionary item was derived using a weakly supervised Naïve Bayes classifier and an RNN/LSTM network. The model is based on the assumption that words with similar abstractness likely share similar contextual usage. The accuracy achieved ranged between 0.65 and 0.74 depending on the model and the manually annotated test set (including the BWS dictionary).

Attribution Methodology

BWK Dictionary Based Attribution

Our primary interest is in exploring the usefulness of abstractness/concreteness as stylistic features for authorship attribution. Our initial approach was based solely on the BWK dictionary. The basic idea was to distribute the words/phrases from each document into categories based on the PoS role of the word and its rating's standard deviation (SD), which represents the agreement of the evaluators as to the word's abstractness. The abstractness of a word differs based on its PoS role in a particular context. For example, the word "horse" is very concrete as a noun (mean: 5, SD: 0.0 in the BWK dictionary) but more abstract as a verb (mean: 2.73, SD: 1.46). Similarly, the noun "orange" is very concrete (mean: 4.66, SD: 0.9), but the color adjective "orange" is fairly abstract (mean: 3.21, SD: 1.32). We used the PoS categories jj, jjr, jjs, nn, nnp, nns, rb, rbr, rbs, vb, vbd, vbg, vbn, vbp, vbz, based on the standard Penn Treebank tags (Santorini, 1990, Penn Treebank Tags). We also added the tag "wp" for word pairs.

We conjectured that the standard deviation of the evaluators' ratings is another indicator of abstractness - the more abstract the term, the larger the disagreement of the evaluators. (Table 1).

Concrete Word	Rating/SD	Abstract Word	Rating/SD
Bread	4.92 / 0.28	Apologize	2.63 /1.62
Balloon	4.92 / 0.27	Forcefully	2.68 /1.42
Doornail	4.90 / 0.40	Infiltrate	2.71 /1.27
Sleeveless	4.60 / 0.50	Modulate	2.52 /1.48
Yogurt	4.90 / 0.31	Venture	2.60 /1.48

Table 1: Ratings/SD of concrete and abstract words

To account for this trait, we defined four SD classes: “very narrow” (SD<0.5), “narrow” (0.5≤SD<1.0), “wide” (1.0≤SD<1.5) and “very wide” (SD≥1.5).

The PoS and SD classes described above were combined to form 64 abstractness classes (e.g., “jj_narrow”, “vb_wide”, etc.). Thus, each text in the corpus is represented by a 64-dimensional vector, where each vector element is the mean abstractness rating of all words which map to that class. The most (universally) common 25 nouns, 50 verbs and their tenses, 50 adjectives, and 35 adverbs are removed to avoid skewing the results by frequently used words. The generated vectors are stored as WEKA (Hall et al 2009) files and used for training the machine learning classifiers, which perform the attribution.

The Multi-Dictionary Approach

The BWK dictionary has a number of flaws and limitations as illustrated by the samples in Table 2 below:

Word	Mean	SD	PoS	Problems
quack	3.93	1.49	Interjection	Can also be noun or verb. As an interjection this is not concrete.
shoeless	4.28	1.10	Name	Name???
tape	4.9	0.31	Noun	Can be a verb but is missing.
drink	4.76	0.69	Noun	Verb is more common but is missing
slit	3.68	1.46	Verb	Noun is more common but is missing
blazing	3.64	1.35	Verb	Morphologically inappropriate as a verb. Can be an adjective but is missing.

Table 2: Some problems with the BWK dictionary

Most of the dictionary entries reflect only a single sense/PoS role of the word. In fact, many words do not have a PoS role indicated. There are many incorrect entries and improper morphological forms. Moreover, despite the authors’ claim that all reviewers were native speakers of English, in many cases the reviewer ratings seem suspect – almost as if the reviewers collectively did not know the meaning of certain words.

Given the limitations of the BWK dictionary, we wanted to add an alternative, independent source of concreteness/abstractness rating. The 300K words dictionary from (Rabinovich et al 2018) was the only other currently available such source. Henceforth, we refer to this dictionary as the IBM/UT dictionary after the affiliations of the authors.

The use of this dictionary is not without its problems. First, the accuracy of the machine generated abstractness/concreteness ratings is fairly low even when tested against the flawed BWK dictionary. Secondly, there is no standard deviation since all ratings were generated by a single machine learning algorithm. Thus, we could not rely on this important abstractness/concreteness trait, nor could we use our original algorithms directly. Indeed, basing our attribution solely on the PoS trait of words yielded significantly lower results compared to those obtained by using the BWS dictionary on the same texts (see Table 3).

We explored several approaches to overcoming this deficiency. In the first approach, the words in a text are looked up in both the BWK and the IBM/UT dictionaries and, if present in both, the average of the word’s abstractness ratings from the two dictionaries is used. The second approach was even simpler – augment the BWS dictionary based 64-dimensional vectors with the 16-dimensional IBM/UT dictionary generated data. As it turned out, the latter approach yielded a marked improvement in the attribution accuracy especially for larger numbers of authors.

Results

We used the Reuters-RCV1 dataset (Lewis et al 2004, NIST) as the training and test corpus. The Reuters corpus includes 50 authors, each represented by a preprocessed collection of 100 texts – 50 for training and 50 for testing. We randomly selected 20-, 15-, and 10 author subsets and used all 50 training texts of each author to train several WEKA (Hall et al 2009) machine learning classifiers using leave-one-out cross-validation. The classifiers used in the experiments included a random forest classifier (RF), a support vector machine with minimal sequential optimization (SMO), and two multilayer perceptrons (MLP). We conducted several sets of experiments varying the classifier parameters. In all experiments, the RF classifier outperformed the SMP and MLP classifiers. Tables 3 below summarizes the results from the RF classifier experiments using the BWK dictionary only, the IBM/UT dictionary only, and the combined multi-dictionary approach:

Num. of Authors / Stylistic Features	20 Authors	15 Authors	10 Authors
Character-2-Grams	69.8%	75.2%	83.6%
Character-3-Grams	59.5%	64.6%	77.4%
M-W function words	45.3%	56.0%	64.7%
First-word-in-sentence	38.0%	47.4%	63.5%
Coarse POS Tagger	54.3%	65.5%	72.8%
Prepositions	38.1%	44.5%	55.3%
Suffices	64.1%	73.1%	80.6%
Vowel-initiated words	49.4%	60.0%	69.4%
Word-2-grams	43.5%	54.4%	66.9%
Abstractness (BWK)	*60.1%*	*71.5%*	*79.0%*
Abstractness (IBM/UT)	*52.6%*	*59.5%*	*64.8%*
Abstractness (Combined)	*65.4%*	*75.7%*	*79.8%*

Table 3: Accuracies from the RF classifier experiments

The results presented in Table 3 indicate that, as a stylistic feature, abstractness/concreteness is among the top three highest accuracy features in the conducted experiments. The BWK-only dictionary approach ranked 3rd in all experiments. The combined multi-dictionary approach was 2nd in the 20-authors experiments, 1st in the 15-author experiments, and a close third in the 10-authors experiments. Notice also that the addition of the IBM/UT dictionary ratings increases the accuracy by as much as 5.3% compared to the BWK dictionary only experiments.

Another interesting observation comes from examining the per-author results as well as the confusion matrices generated in the WEKA experiments. The precision, recall, and F-measure from Table 4 are fairly high for most authors. However, some authors exhibit especially high precision and recall (e.g., Fumiko Fujisaki: 0.920/0.876 Lydia Zajc: 0.878/0.860, Graham Earnshaw 0.860/0.860, etc.). This remains true when these authors are paired with other randomly selected authors from the Reuters RCV-1 dataset and regardless of the type of classifiers used. The observation is confirmed by examining the confusion matrix in Fig. 1: Lydia Zajc, Graham Earnshaw, Fumico Fujisaki, and Peter Humphrey have the highest number of correctly attributed papers (43/50, 43/50, 46/50, and 47/50 respectively). This may imply that some authors have a more abstract or concrete writing style, using abstract/concrete words and phrases in unique ways that makes their writing style more easily identifiable.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.760	0.040	0.679	0.760	0.717	0.685	0.952	0.822	MarcelMichelson
0.700	0.022	0.778	0.700	0.737	0.711	0.940	0.788	DarrenSchuettler
0.860	0.013	0.878	0.860	0.869	0.854	0.954	0.877	LydiaZajc
0.780	0.027	0.765	0.780	0.772	0.747	0.966	0.864	DavidLawder
0.860	0.016	0.860	0.860	0.860	0.844	0.976	0.902	GrahamEarnshaw
0.780	0.016	0.848	0.780	0.813	0.793	0.975	0.874	RogerFillion
0.920	0.020	0.836	0.920	0.876	0.863	0.994	0.963	FumikoFujisaki
0.680	0.016	0.829	0.680	0.747	0.727	0.958	0.860	BernardHickey
0.700	0.016	0.833	0.700	0.761	0.740	0.956	0.838	ToddNissen
0.940	0.040	0.723	0.940	0.817	0.803	0.987	0.899	PeterHumphrey
0.798	0.022	0.803	0.798	0.797	0.777	0.966	0.869	

Table 4: Per-author TP rate, precision, recall, and F-measure

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i  j   <-- classified as
 38  2  1  2  1  1  0  2  1  2 |  a = MarcelMichelson
  2 35  3  0  0  2  3  2  0  3 |  b = DarrenSchuettler
  1  1 43  0  1  1  1  0  1  1 |  c = LydiaZajc
  1  0  0 39  0  1  2  1  3  3 |  d = DavidLawder
  1  1  0  1 43  0  0  1  0  3 |  e = GrahamEarnshaw
  1  1  1  2  1 39  1  0  1  3 |  f = RogerFillion
  0  1  0  0  1  0 46  1  0  1 |  g = FumikoFujisaki
  7  2  1  0  2  1  2 34  0  1 |  h = BernardHickey
  4  2  0  7  0  1  0  0 35  1 |  i = ToddNissen
  1  0  0  0  1  0  0  0  1 47 |  j = PeterHumphrey
```

Fig.1: Confusion Matrix from a 10-author experiment

Conclusion and Future Work

The work presented in this paper is the beginning of a much more in-depth investigation into the usefulness of abstractness/concreteness as stylistic features for authorship attribution. The results obtained thus far appear very promising, but we need to conduct further experiments with additional corpora. At present we are experimenting with a 20-author corpus of 19th century literary works extracted from project Gutenberg. We are also preparing a corpus of 19th and 20th century poetry. The results from poetry experiments should be particularly telling since poetry naturally involves the wide use of abstraction.

Additionally, our methodology requires further attention: The most important – and time consuming – task will be to correct the numerous issues in the BWK dictionary. With the increased interest in the computational aspects of abstractness/concreteness, we hope that a better abstractness/concreteness dictionary will emerge, but until then, the BWK dictionary remains the primary resource for abstractness/concreteness researchers. While we cannot add the missing ratings based on the different PoS roles of the same word, we can correct the incorrect PoS tags as well as the morphological issues we have encountered. We also need to consider the use of word embeddings since the abstractness/concreteness of words is very much dependent on the context. Additional research which may improve the performance of our methodology includes differentiating abstractness ratings by "generality" vs. "sensory perceptibility" as well as addressing the issue of abstractness/concreteness complementarity and its impact on authorship attribution. An interesting, but challenging research direction is the study of metaphors and hyperbole as special cases of abstractness and their potential role in authorship attribution. It may also be worth exploring new (semi-)automatic methods for rating abstractness/concreteness – an extension of the efforts of Köper and Schulte im Walde, and Maudslay and Rabinovich and their teams. While challenging, any of these investigations hold the promise of further improving the performance of our methodology and adding another reliable and high-performing stylistic feature to the set of standard features used in modern authorship attribution.

References

Assaf D., Neuman Y., Cohen Y., Argamon S., Howard N., Last M., Frieder O., Koppel M. 2013. Why "dark thoughts" aren't really dark: A novel algorithm for metaphor identification. In *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2013 IEEE Symposium on*, pages 60–65. IEEE

Birke J., Sarkar A. 2006. A Clustering Approach for the Nearly Unsupervised Recognition of Nonliteral Language. In *Proceedings of the 11th Conference of the European Chapter of the ACL*, pages 329–336, Trento, Italy.

Birke J., Sarkar A. 2007. Active Learning for the Identification of Nonliteral Language. *Proc. of Workshop on Computational Approaches to Figurative Language*, pages 21–28, Rochester, NY.

Burgoon E., Henderson M., Markman A. 2013. There Are Many Ways to See the Forest for the Trees: A Tour Guide for Abstraction, *Perspectives on Psychological Science* 2013 8: 501 DOI: 10.1177/1745691613497964

Brysbaert, M., Warriner, A., Kuperman, V. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46, 904-911.

Coltheart, M. 1981. The MRC psycholinguistic database. *The Quarterly J. of Experimental Psychology,* 33, 497-505.

Gao G., Choi E., Choi Y., Zettlemoyer L. 2018. Neural metaphor detection in context. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 607–613, Brussels, Belgium. Assoc. for Computational Linguistics.

Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I., 2009. *The WEKA Data Mining Software: An Update; SIGKDD Explorations*, Volume 11, Issue 1

Ivanov, L. 2019. Learning Patterns of Assonance for Authorship Attribution of Historical Texts: FLAIRS-32, Sarasota, FL, USA., 5/19, pp. 191-196.

Ivanov, L., Aebig A., Meerman S. 2018. Lexical Stress-Based Authorship Attribution with Accurate Pronunciation Patterns Selection: TSD 2018, Brno, Czech Rep., 9/18, pp. 67-75.

Ivanov, L. 2016. Using Alliteration in Authorship Attribution of Historical Texts, Text, Speech, and Dialogue. TSD 2016. Lecture Notes in Computer Science, vol. 9924. Springer

Ivanov, L., Petrovic, S. 2015. Using Lexical Stress in Authorship Attribution of Historical Texts, Chapter, LNCS: TSD, v.9302, pp.105-113

Juola P. 2009. JGAAP: A System for Comparative Evaluation of Authorship Attribution. Journal of Chicago Colloquium on Digital Humanities and Comp. Science, 1(1): 1-5.

Klebanov B., Leong C., Flor M. 2015. Supervised word-level metaphor detection: Experiments with concreteness and reweighting of examples. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 11–20, Denver, CO. ACL.

Köper, M & Schulte im Walde, S. 2016. Automatically Generated Affective Norms of Abstractness, Arousal, Imageability and Valence for 350 000 German Lemmas. *LREC'16*.

Kwong O.Y. 2011, Measuring Concept Concreteness from the Lexicographic Perspective, *25th Pacific Asia Conference on Language, Information and Computation, pages 60–69*

Lewis, D. D.; Yang, Y.; Rose, T.; and Li, F. 2004. RCV1: A New Benchmark Collection for Text Categorization Research. *Journal of Machine Learning Research*, 5:361-397

Maudslay R., Pimentel T., Cotterell R., Teufel S., 2020. Metaphor Detection using Context and Concreteness, *Proceedings of the Second Workshop on Figurative Language Processing*, pages 221—226.

Neuman Y., Assaf D, Cohen Y., Last M., Argamon S., Howard N., Frieder O. 2013. Metaphor identification in large texts corpora. *PloS One*, 8(4):e62343.

NIST: https://trec.nist.gov/data/reuters/reuters.html

Paivio, A., Yuille. J., Madigan, S., 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology*, 76(1p2):1

Penn Treebank Tags: https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

Petrovic I., Petrovic S., Palesi I., Calise A. 2019. Eliminating Sycophants to Improve Authorship Attribution, *Proceedings of FLAIRS-32 Conference*.

Petrovic S., Petrovic I., Palesi I., Calise A., 2018. Weighted Voting and Meta-Learning for Combining Authorship Attribution Methods, *Proceedings of Intelligent Data Engineering and Automated Learning–IDEAL 2018*: 19th International Conference, Madrid, Spain, Part I 19, pages 328-335

Rabinovich E., Sznajder B., Spector A., Shanayderman I., Aharonov R., Konopnicki D., Slonim N. 2018. "Learning Concept Abstractness Using Weak Supervision", *proceedings of EMNLP 2018*, pages 4854-4859, Brussels, Belgium

Rei M., Bulat L., Kiela D., Shutova E. 2017. Grasping the finer point: A supervised similarity network for metaphor detection. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1537–1546.

Santorini B., 1990 "Part-of-Speech Tagging Guidelines for the Penn Treebank Project (3rd Revision)"

Shutova E., Teufel S., Korhonen A. 2013. Statistical Metaphor Processing. *Computational Linguistics*, 39(2):301–353.

Sporleder C., Li L. 2009. Unsupervised Recognition of Literal and Non-Literal Use of Idiomatic Expressions. In *Proceedings of the 12th Conference of the European Chapter of the ACL*, pages 754–762, Athens, Greece.

Spreen O., Schulz. R. 1966. Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning and Verbal Behavior*, 5(5):459–468.

Stamatatos E. 2016. Authorship Verification: A Review of Recent Advances, Research in Computer Science, 123, pp.9-25,IPN.

Stamatatos E. 2009. A Survey of Modern Authorship Attribution Methods, Journal of the American Society for Information Science and Technology, 60(3), pp. 538-556, Wiley

Tsvetkov Y., Boytsov L., Gershman A., Nyberg E., Dyer C. 2014. Metaphor Detection with Cross-Lingual Model Transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 248–258.

Tsvetkov Y., Mukomel E., Gershman A. 2013. Cross-lingual metaphor detection using common semantic features. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 45– 51, Atlanta, Georgia. ACL.

Turney P., Neuman Y., Assaf D., Cohen Y. 2011. Literal and Metaphorical Sense Identification through Concrete and Abstract Context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 680–690, Edinburgh, UK.

Wu C., Wu F., Chen Y., Wu S., Yuan Z., Huang Y. 2018. Neural metaphor detecting with CNN-LSTM model. In *Proceedings of the Workshop on Figurative Language Processing*, pages 110–114, New Orleans, Louisiana. ACL.