Do Idioms Have a Heart? The SIDE (Sentiment of IDiomatic Expressions) Project

Lucia C. Passaro[1], Marco S. G. Senaldi[2], and Alessandro Lenci[1]

1 – University of Pisa, Italy
2 – Scuola Normale Superiore, Pisa, Italy
lucia.passaro@fileli.unipi.it, marco.senaldi@sns.it, alessandro.lenci@unipi.it

ABSTRACT

This paper presents some preliminary results of the SIDE project, which aims at investigating the emotional content of idioms from both a behavioral and computational point of view. In this first work, we collected affective ratings for a set of 45 Italian verb-noun idioms and 45 Italian non-idiomatic verb-noun pairs and then performed a computational study to analyze the role of linguistic information in determining the affective rating of idioms and non-idioms. To this end, we exploited an existing vector-based computational model to calculate a distributional polarity score. The correlation analysis between such a score and human-rated valence revealed a high correlation among the variables as well as a significantly lower polarity score of idioms with respect to non-idioms reflecting their stronger negative content.

Keywords: idiomatic expressions, emotive content, Distributional Semantic Models

1. INTRODUCTION

An idiomatic sentence like *John has kicked the bucket* does not just mean that *John has died*, but it also expresses this concept in a quite informal manner, which could differ from the literal equivalent for its valence and arousal. Warriner and colleagues [27] define these dimensions as follows: Valence is the pleasantness of the stimulus, usually measured on a scale from 1 (very unpleasant) to 9 (very pleasant). Arousal is instead the intensity of the feeling evoked, on a scale from "stimulated" to "unaroused". In other words, valence indicates the degree of positivity/negativity of the conveyed concept and arousal says how emotionally intense the described event is [14, 19]. A key aspect of idiomatic expressions (e.g., *kick the bucket* 'to die', *get the sack* 'to be fired', etc.) that has been mostly left aside in previous literature is their affective value [19]: in addition to being semantically non-compositional and formally rigid [3], idioms are more emotionally loaded than literal expressions [8, 7] and add some rhetoric and affective nuances to their meanings which are worth investigating to fully understand their semantics and their usage in language. The contribution of the SIDE project is to combine behavioral and computational evidence on idiom sentiment and emotional content. In this first work, we elicited affective ratings for a set of 45 Italian verb-noun idioms and 45 Italian non-idiomatic verb-noun pairs and we compared them with polarity scores obtained via distributional semantics models.

The paper is organized as follows: Section 2 describes related work and resources, Section 3 shows the methodology employed in this study to collect both behavioural and distributional ratings and Section 4 reports on the results.

2. RELATED WORK

2.1. Figurative expressions and emotions

Previous literature showed that figurative language is extensively used for emotion expressions [11, 12, 13] and when recalling emotional past events [8] and that it is especially useful in emotional communication [9]. fMRI studies have actually confirmed that metaphors activate brain areas related to emotion processing [6]. As for idioms specifically, discourse analytic research has shown speakers to prefer them when manifesting complaints [7], probably by virtue of their indirectness and in order to elicit empathy from the interlocutor. While a normative study on German idioms found idioms with negative valence to be more prevalent than those with positive valence, a U-shaped quadratic relationship between valence and arousal was observed, whereby more arousing idioms tend to be more positively or negatively valenced [5].

The work we report on has resorted to two main existing emotive resources, namely the Italian version of the Affective Norms for English Words (ANEW) [18] and the Italian EMotive lexicon (ItEM) [22, 21].

2.2. Emotive resources

The Italian adaptation [18] of ANEW [2] contains the norms for the translation of the original ANEW words, as well as for words taken from the Italian Semantic Norms [17]. The three main dimensions of valence, arousal and dominance were rated using a 1 to 9 scale (rate 1 for valence means unpleasant and 9 means very pleasant), in order to provide consistency with the original norms. Apart from the original affective ratings, new dimensions were collected as well, namely subjective and objective psycholinguistic indices. The Italian EMotive lexicon (ItEM [21,22]) is a distributional resource based on the Distributional Hypothesis [10], which states that semantically similar words tend to appear in similar contexts. In ItEM, this hypothesis has been generalized to emotions stating that "a word *w* is associated with an emotion *E* if it co-occurs in similar contexts of other words associated with *E*" [21]. This means that by constructing a centroid vector starting from a set of words strongly associated to a particular emotion (the seed words), it is possible to approximate the emotive content of all unknown words starting from their distributional profile. To implement this hypothesis, each emotion in ItEM has been represented as a centroid vector built out of a set of seed words strongly associated to each of the eight Plutchik's basic emotions (JOY, SADNESS, ANGER, FEAR, TRUST, DISGUST, SURPRISE, ANTICIPATION). The vectors were originally built on a classical count-based distributional model [22] and then adapted to a prediction-based one [20].

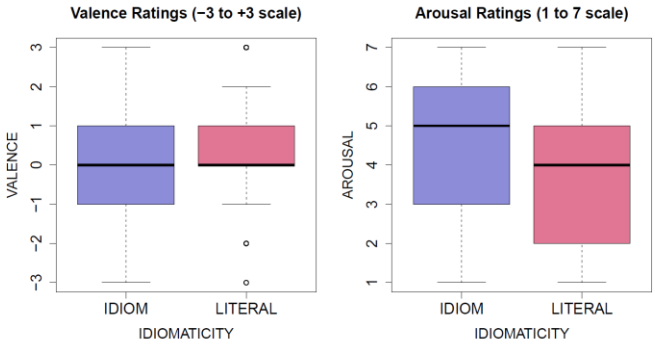
3. METHOD

In order to study the relationship between human-elicited emotional scores and distributional ones, we crowdsourced behavioural data on a sample of 90 target expressions including 45 Italian verb-noun idioms and other 45 Italian non-idiomatic literal verb-noun phrases that roughly matched the frequency range of the 45 idioms and were selected from an online resource on Italian verb distributional profiles [15]. More specifically, participants to two questionnaires posted on the platform Figure Eight rated our targets for valence on a -3 to +3 scale (24 subjects) and for arousal on a 1-7 scale (25 subjects). The -3 to +3 scale, which reflects the bipolar nature of valence, was chosen in line with previous normative studies [5, 25]. A Wilcoxon test on the behavioural data showed that idioms were rated as significantly more negative than non-idioms and, as expected, as significantly more arousing.

Table 1: Results of the Wilcoxon test on Valence and Arousal human-elicited ratings (675 data points).

	Mean idioms	Mean literals	Wilcoxon
Valence	0.20	0.51	-3.19 (p<.01)
Arousal	4.37	3.48	9.62 (p<.001)

Figure 1: Boxplots for Valence and Arousal human ratings on the 45 idioms and 45 literals.



The greater negativity exhibited by idioms had already been brought to light in previous normative [5] and discourse-analytic [7] studies.

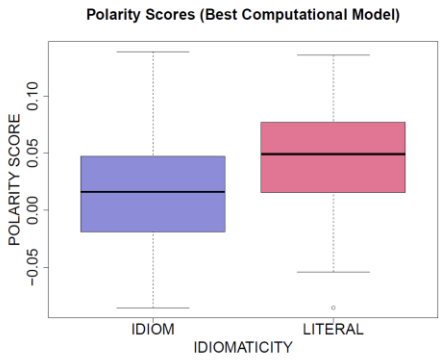
Simultaneously, we carried out a computational study to analyze the role of linguistic information in determining the affective rating of idioms and non-idioms. We elaborated on ItEM [20], from which we borrowed both the algorithm to infer the emotional content of the expressions and the seed words. We explored various parameter settings including (i) vector type (count- or prediction-based); (ii) the seed words used to construct the centroid vectors (taken from ItEM or ANEW) and (iii) the dimensionality of the built vectors. As for the distributional model, we extended the model for single-word emotion detection in [22] and we computed the valence for our idiomatic and literal phrases by encoding them as single tokens, without composing the vectors of their subparts. To do so, we pre-processed the itWaC corpus [1] by concatenating the phrase components and placing potential intervening material right after the expression. For instance, the sentence *tirare velocemente le cuoia* ('to quickly kick the bucket') was transformed into *tirare_le_cuoia velocemente*. Both count [26] and prediction (i.e., word2vec [16]) vector representations were built for the 90 target phrases from the itWaC corpus [1]. For count-based DSMs, we used as target and contexts the 30,000 most frequent nouns, verbs and adjectives as well as our 45 idiomatic and 45 non-idiomatic phrases taken as single tokens. In such models the context has been defined as the entire sentence and raw co-occurrences

were weighted with Positive Pointwise Mutual Information (PPMI) [4]. Vectors were then reduced with Singular Value Decomposition (SVD) to 300, 500 and 1000 dimensions. Prediction-based DSMs were built by means of Skip-Gram with Negative Sampling algorithm. The context span was set to 5; the occurrence threshold was set to $1 * e{-}4$, and the number of negative examples was set to 10. We trained vectors with 300, 500 and 1000 dimensions. Cosine similarity was finally measured between the vector of each target and the centroids representing positive and negative polarity. For what concerns the seed words, we tried with both ANEW and ItEM ones. In the former case (i) we simply applied a threshold on the arousal (≥ 5, ≥ 6 and ≥ 7) and we restricted the analysis to the words belonging to our distributional space (i.e., the 30,000 most frequent nouns, verbs and adjectives). In the latter case (ii) we mapped the eight Plutchik's basic emotions to two polarity centroid vectors. In particular, the seeds elicited in [22] just for the emotions JOY and TRUST have been grouped into the POSITIVE class and the seeds elicited just for SADNESS, ANGER, FEAR and DISGUST have been classified as NEGATIVE. The emotions SURPRISE and ANTICIPATION have been left out because of their mixed nature, since they can be interpreted as referring to both positive and negative events [22]. Starting from the polarity values calculated with the cosine between the target phrase vectors and the two centroids, a unique polarity score was obtained by: i) picking the greater value between the cosine similarity of a target with the POSITIVE centroid (POS) and its similarity to the NEGATIVE centroid (NEG) and switching the sign to negative in the latter case (MAX models); ii) subtracting NEG from POS (POS-NEG models).

4. RESULTS AND DISCUSSION

A correlation analysis was run between our computational polarity scores and human-elicited valence ratings. Table 2 shows the correlation results obtained in the top 5 models. All models use prediction vectors and the POS-NEG function to calculate the polarity score. It is clear that such function is just an approximation of valence, since it tends to provide words with both a high positive and high negative score with a neutral one, but on average it seems in line with human ratings. The POS-NEG model with 300-dimension prediction vectors and ItEM-extracted seeds (variant 1) appears to perform the best (ρ=0.60***). As with human ratings, idiom polarity scores were significantly lower than those given to non-idioms (W=-2.51, $p < .05$), reflecting once again their stronger negative content.

Figure 2: Boxplot for the polarity scores computed by the best model for the 45 idioms and 45 literals.



An even higher correlation between the scores and the ratings was observed when keeping just the 23 idioms and 19 non-idioms with an arousal value higher than the mean of each class (ρ=0.70***). Such an improvement indicates that distributional polarity scores are more effective on more arousing words. Given the quadratic relationship [2] between valence and arousal, we could argue that the method works better for very positive and very negative expressions.

Table 2: Best 5 computational models in terms of correlation with the Valence and Arousal human ratings.

Seeds	Seeds number (pos+neg)	Vector dims	Spearman's ρ
ItEM	**84+149**	**300**	**0.60*** **
ANEW (aro > 5)	210+259	500	0.56***
ANEW (aro > 6)	122+188	300	0.56***
ANEW (aro > 5)	210+259	300	0.55***
ANEW (aro > 5)	210+259	1000	0.54***

Note. *** = $p < .001$.

5. CONCLUSION AND FUTURE DIRECTIONS

We know from previous literature that idioms are preferred by speakers when recounting emotional events or to manifest complaints, maybe by virtue of their indirectness [8, 7]. This study confirmed that idioms are more emotionally loaded and negative than literals [5]. Distributional data have been proven effective for single-word emotion detection and seem to be able to model idiom affect given that DSM-based polarity scores tally with speaker-elicited ratings on valence and arousal. This study just aimed at a preliminary investigation on idiom sentiment. A comparison between the affective

value of idioms (*kick the bucket*) and that of their respective meanings (*to die*) is lacking. Moreover, the relationship between semantic compositionality and sentiment composition could be a further key element in determining the affective rating of idioms. Future research will then be addressed to: (i) extending the dataset to new idioms and their respective paraphrases and (ii) carrying out computational studies on the relationship between the holistic sentiment of idioms and the one associated with their components.

REFERENCES

1. Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. Language resources and evaluation, 43(3), 209-226.
2. Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings (Vol. 30, No. 1, pp. 25-36). Technical report C-1, the center for research in psychophysiology, University of Florida.
3. Cacciari, C. (2014). Processing multiword idiomatic strings: Many words in one? The Mental Lexicon, 9(2), 267-293.
4. Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. Computational linguistics, 16(1), 22-29.
5. Citron, F. M., Cacciari, C., Kucharski, M., Beck, L., Conrad, M., & Jacobs, A. M. (2016). When emotions are expressed figuratively: Psycholinguistic and Affective Norms of 619 Idioms for German (PANIG). Behavior research methods, 48(1), 91-111.
6. Citron, F. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. Journal of cognitive neuroscience, 26(11), 2585-2595.
7. Drew, P., & Holt, E. (1988). Complainable matters: The use of idiomatic expressions in making complaints. Social problems, 35(4), 398-417.
8. Fainsilber, L., & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. Metaphor and Symbol, 2(4), 239-250.
9. Gibbs Jr, R. W., Leggitt, J. S., & Turner, E. A. (2002). What's special about figurative language in emotional communication? In The verbal communication of emotions (pp. 133-158). Psychology Press.
10. Harris, Z. S. (1954). Distributional structure. Word, 10(2-3), 146-162.
11. Kövecses, Z. (2003). Metaphor and emotion: Language, culture, and body in human feeling. Cambridge University Press.
12. Kövecses, Z. (2012). Emotion concepts. Springer Science & Business Media.
13. Lakoff, G., & Johnson, M. (2008). Metaphors we live by. University of Chicago press.
14. Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. NIMH Center for the Study of Emotion and Attention, 39-58.
15. Lenci, A., Lapesa, G., & Bonansinga, G. (2012). LexIt: A Computational Resource on Italian Argument Structure. LREC 2012 (pp. 3712-3718).
16. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).
17. Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). Semantic memory: A feature-based analysis and new norms for Italian. Behavior research methods, 45(2), 440-461.
18. Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of the affective norms for English words (ANEW) for Italian. Behavior research methods, 46(3), 887-903.
19. Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. Language, 70(3), 491-538.
20. Passaro L. C., Bondielli A. and Lenci A. (2017). Learning Affect with Distributional Semantic Models. Italian Journal of Computational Linguistics 3(2), 23-26.
21. Passaro, L. C. and Lenci, A. 2016. Evaluating context selection strategies to build emotive vector space models. LREC 2016 (pp. 2185-2191).
22. Passaro, L., Pollacci, L., & Lenci, A. (2015). ItEM: A vector space model to bootstrap an italian emotive lexicon. In Second Italian Conference on Computational Linguistics CLiC-it 2015 (pp. 215-220). Academia University Press.
23. Plutchik, R. (1994). The psychology and biology of emotion. New York, NY, US: HarperCollins College Publishers.
24. Russell, J. A. (2003). Core affect and the psychological construction of emotion. Psychological review, 110(1), 145.
25. Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: Affective norms for German sentiment terms, derived from the affective norms for English words. Behavior research methods, 46(4), 1108-1118.
26. Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. Journal of artificial intelligence research, 37, 141-188.
27. Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior research methods, 45(4), 1191-1207.